



Recent changes in the global rail industry: facing the challenge of increased flexibility

Pedro Cantos^{1*}, Javier Campos²

¹ *Departamento de Análisis Económico (Universidad de Valencia)*

² *Departamento de Análisis Económico Aplicado (Universidad de Las Palmas de Gran Canaria)*

Abstract

This paper discusses how the current trend towards increased private participation in the rail industry is reshaping the way in which Governments should address the main regulatory challenges arising from the particular economic and technical characteristics of this industry. We review the role of railroads in the last two decades and examine the characteristics of the most relevant processes of private participation around the world. The lessons learned from these changes suggest that many of the traditional regulatory paradigms in this industry are being replaced by more flexible schemes of public intervention. Although this change does not fully preclude direct participation by the Government, it seems that the traditional monopolistic rail company is dead as the dominant model around the world, and new forms, such as franchises or concessions competing on the tracks are progressively gaining relevance.

Keywords: Railways; Restructuring; Regulation; Privatization.

1. What makes rail regulation different?

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. These characteristics define a descriptive framework for this sector, and jointly determine the main factors that should be considered when studying in detail the appropriate economic regulation for the rail industry.

* Corresponding author: Pedro Cantos. Departamento de Análisis Económico (Universidad de Valencia). Campus dels Tarongers, s/n. 46022 Valencia (Spain). E-mail: Pedro.Cantos@uv.es.

1.1. The multi-product nature of the rail activity

Rail companies are, in most cases, multi-product firms that provide different types of freight and passenger transport services. In the case of freight, along with the usual transport of bulk freight, rail operators also supply complete cargo wagons or trains, parcel and postal services, as well as other services of intermodal transport. In the case of passenger transport, long-distance traffic usually co-exists with local services (suburban and commuter trains), regional services, and in certain cases, even with high-speed trains.

The multi-product nature of railways has different implications. In accounting, for example, it is often difficult to allocate total operating costs among services. Many of the costs of running a long-distance train (including not only infrastructure costs but also variable costs) are shared by different types of traffic and these joint costs co-exist with other costs not affected by changes in output. For instance, the common costs of signal maintenance along a line section usually do not increase if the proportions of traffic of the different services change. Although some cost elements may be attributable to a particular traffic (for example, passengers), most of them (wagons, energy, staff, etc.) are not. Thus, cost interdependence requires simultaneous decisions on prices and services, which, in practice, makes any regulatory task much harder.

At the cost level, another important aspect to consider in the multi-product setup of the rail industry is the sub-additivity of the cost function faced by a railroad.¹ This idea conveys two relevant implications for the rail industry. First, is it more efficient for a single firm, rather than two separate firms, to supply both infrastructure and transport services? Second, if the infrastructure and services are separated, is the supply of such services more efficient within the context of a monopoly, or should two or more firms participate. This analysis, connected to the advantages and disadvantages of the separation of infrastructure from services, will be discussed in depth.

1.2. The pervasive structure of railway costs

Waters (1985) broadly distinguishes four railway cost categories: (i) *train working costs*, including the cost of providing transport services (fuel, crew, maintenance and depreciation of rolling stock); (ii) *track and signalling costs* (including operation, maintenance and depreciation of infrastructures); (iii) *terminal and station costs*; and finally, (iv), *administration costs*.

The first two categories are prevalent in most companies and change according to several factors. Among train working costs, for example, rolling stock costs depend on both their number and the distance they run. Fuel costs depend on car-kilometres run for each type of vehicle, while train crew costs vary according to train-kilometres run. Track and signalling costs usually rely on the length of the route (since they typically request a single, standard-quality track). The amount of track and signalling needed, however, changes with the number of trains requiring paths, although this relationship is not constant. Terminal and station costs depend on traffic volumes, but vary considerably with the type of traffic. For instance, bulk freight handling requires more

¹ According to Baumol (1977), a cost function is sub-additive when the provision of services by a single firm is more efficient (in terms of a lower unit cost) than the same production carried out by two or more companies.

terminal expenses than parcel services. Similarly, long distance passengers require more services (ticketing, reservations, luggage, etc.) than short distance users. Administration costs, finally, fluctuate depending on the overall size of the firm, although the precise nature of this dependence is generally difficult to determine.

Allocating all of these costs to the multiple outputs or inputs it produces is complex. It often involves a degree of arbitrariness that demands, from a regulatory point of view, a clear distinction between avoidable and unavoidable costs. The avoidable costs are uniquely associated with a particular output: were this output not produced, no cost would be incurred. Avoidable costs may therefore be considered as a regulatory price floor (if any), since charging less would be equivalent to operating at an economic loss.

1.3. The economic role of rail infrastructure

Since the birth of the rail industry in the last century, mainstream economists have always considered that the larger the size of a railway company, the greater its efficiency. The existence of substantial fixed costs (particularly, those associated with infrastructure) traditionally led economists to assume the presence of important economies of scale, and thus to regard rail transport service as a textbook example of a natural monopoly.

However, this notion has been heavily challenged in recent decades by the introduction of new ideas into the industry's economic analysis. Particularly, the upheaval of the theory of contestable markets (Baumol, Panzar and Willig, 1982) contributed to clarifying the proper definition of the natural monopoly concept, in terms of the sub-additive cost function (see note 1). This concept implies that duplicating rail infrastructure is generally inefficient (and therefore is subject to natural monopoly conditions), but once the network has been deployed, the cost of operating rail transport services and rolling stock can be efficiently covered by more than one company, either as actual or potential competitors.

Therefore, from the regulatory point of view, the conclusion is that infrastructure and services can be dealt with in different ways: the former, as a natural monopoly,² but also as a potential provider of adequate access to any willing-to-serve operator; the latter, as any other competitive economic activity that could be provided by multiple competing operators or by a single firm under some sort of concession or license arrangement.

1.4. The implications of asset indivisibilities

Even though this potential vertical separation alleviates some of the natural monopoly problems, the rail industry remains very capital-intensive, with several other indivisibilities within its productive process. Specifically, the capital units (rolling stock, tracks and stations) can only be expanded in discrete, indivisible increments (the addition of a train or wagon, for example), while demand fluctuates in much smaller units. Consequently, increases (decreases) in supply can exceed increases (decreases) in demand, resulting in excess capacity. This lumpiness has several important implications for investment and pricing. For example, the transportation costs of an additional unit of

² At least, when the infrastructure has not yet been built, although not necessarily after that moment.

traffic (freight or passengers) may be insignificant when there is idle capacity, but may become substantial when the capital is being used to its fullest.

Firms can also be forced to employ fixed assets with differing economic lives, whose reliability spans over a large time horizon and heterogeneously affects the cost items described above, modifying investment decisions, and requiring a complete accounting and management information system. Therefore, dynamic price and output considerations become crucial in order to recover the real costs associated with each period of activity.

A final implication of the indivisibilities in the rail industry's capital assets is that innovation and infrastructure improvement projects are usually deferred and only carried out in small, discrete amounts. Railway firms seldom change the entire definition of their existing network, which in most countries corresponds to an inherited burden from past decades when the traffic structure was very different than today. Instead, they opt for partial renovations that often introduce technical asymmetries between tracks within a country or region, and accentuate indivisibilities and inflexibilities.

1.5. The role of rail transport as a public service

Although not derived from historical and organizational reasons and not from technical characteristics, the concept of rail transportation as a public or social service, irrespective of profitability, is another defining element that has determined the industry's organization and performance around the world. The low rolling resistance of steel wheels on steel rails made railroad transportation extremely fuel efficient and relatively cheap. This allowed railroads to rapidly grow as the first mass transportation system, particularly for passengers, beginning in the years of the industrial revolution.

For military and industrial reasons, some form of public control was envisaged in most countries, and many imposed their control by legal mandate. Public control over the rail industry occurred both with and without accompanying subsidies, public service obligations to transport providers in the form of compulsory (often unprofitable) routes, organized timetables or particular services for strategic products or areas. The ultimate reason behind this control, which remains the same today, is that this industry is regarded as an integral mechanism to overcome geographical barriers in certain areas, aid in the economic development of undeveloped zones, and even as a guarantee of minimum transport services for a particular segment of the population.

1.6. Externalities and the rail system

The policy goal of public service obligation is often supported by the idea that rail transportation contributes less to negative externalities than other transport modes, especially roads. There is abundant empirical evidence showing that under high demand conditions, the external costs of traffic congestion, accidents and environmental impact (noise, visual impact, pollution, etc.) could be reduced by transferring a substantial part of road traffic to rail.

The current intermodal misallocation (more road users than rail users) arises from the fact that road transport does not fully internalize all of the social costs that it generates.

Economists often recommend the use of congestion and/or pollution rates to account for this. However, when these mechanisms are not feasible or politically viable, it might be preferable to decrease railway fares to improve the overall intermodal balance, which is an additional consideration for rail regulation.

In summary, all of the above-mentioned characteristics suggest that an analysis of the regulation of railway transport should be carried out within a general context, taking into account the industry's technological and organizational features, beginning with a detailed evaluation of recent performance.

2. Recent regulatory trends in the global rail industry

The overall evolution of rail transportation in recent years as compared to other transport modes is summarized in Table 1 for OECD countries. There was a substantial fall of the market share in both freight and passenger markets during the 1970s and 1980s, which stabilized (even with a slight increase) during the last five years. In relative terms the decline is particularly relevant because it was during a period when the total volume in both markets grew about 50%, implying that the railroads were not able to take advantage of growing demand in these years.

Table 1: Market shares of different transport modes (1970-2002).

	<i>Passenger traffic</i>						
	1970	1980	1985	1991	1994	1998	2002
Rail (%)	10.43	8.64	7.33	6.92	6.85	6.83	7.04
Private car (%)	77.30	79.97	83.37	84.37	84.38	84.48	84.64
Bus (%)	12.26	11.38	9.29	8.70	8.75	8.68	8.30
	<i>Freight traffic</i>						
	1970	1980	1985	1991	1994	1998	2002
Rail (%)	31.3	23.2	21.2	17.9	15.5	14.3	15.0
Road (%)	55.2	65.9	69.3	74.0	76.2	78.5	77.6
Waterways (%)	13.5	10.9	9.5	8.1	7.9	7.2	7.4

Source: CEMT. *Evolution des Transports. OECD Countries.*

The substantial reduction in market share is not particular to OECD countries but a common trend around the world. It can be attributed to both exogenous and endogenous causes. The former includes the rapid development of alternative modes of transport, especially road. For passengers, economic growth fostered the development of the automobile market, leading to enormous growth in motorization. In freight transport, the expanding, competitive trucking sector gained a growing percentage of transport in many countries. For example, in 1970 in Europe, there were 150 cars per 1,000 inhabitants, a figure that now is 424. Similarly, the number of heavy vehicles and trucks increased from 7 to 20 million from 1970-2000.

The endogenous causes of the decline can be summarized in the inability of the sector to adapt to the changing conditions of the economic environment. Regulation remained obsolete and the rail industry was slow to react. The policies adopted during the 1980s, as described below, did not halt the steady loss of market share, the growing financial deficits, and in some countries, the impossibility of raising the low productivity indices of the industry. Thus, more radical restructuring processes were put into practice.

2.1. The traditional railroad model

During the past fifty years, the most common market structure in many countries' rail sectors was a single, state-owned firm, entrusted with the unified management of both infrastructure and services. Despite some differences in their degree of commercial autonomy, the traditional methods of regulation and control of this sort of company have been relatively homogeneous. In general, it was assumed that the monopoly power of the national company required price and service regulation to protect the general interest. In addition, there was an obligation on the part of the companies to meet any demand at those prices. The closure of existing lines or the opening of new services required government approval. Thus, competition was rare and often discouraged, and the preservation of the national character of the industry was considered the key factor governing the overall regulatory system.

Under this protective environment, most national rail companies incurred growing trade deficits during the 1970s and 1980s. Furthermore, social obligations to their staff made it nearly impossible to reach any agreement on redundancies or even wage adjustments. In some countries, the companies were forced to finance their deficits by borrowing, so their accounts lost all resemblance to reality. The main problems associated with the traditional policies for railways were: *(i)* increasing losses, which were usually financed by public subsidies; *(ii)* a high degree of managerial inefficiency; and *(iii)* business activities oriented exclusively toward production targets rather than commercial and market targets.³

These distortions did not come from any artificial reduction in the range of services provided, nor from excessively high fares, but more commonly, from an unjustified increase in the supply of services (and hence, of costs). Such behaviour implied larger public subsidies. In many cases, the lack of commercially-oriented tariffs and investment policies explained many of the difficulties faced by the companies. Together with the burden imposed by the technical characteristics of the sector, this placed most railways in a very weak position to compete against alternative transport modes. However, fierce intermodal competition was not able to improve the competitiveness of the railway system by itself. It was necessary to adopt measures affecting the internal behaviour and structure of the sector itself. Therefore, the sector's overall decline sparked a widespread restructuring movement around the world.

³ On this point, Oum and Yu (1994) and Gathon and Pestieau (1995) have empirically shown that the companies that achieve the greatest efficiency were those that had been run with a higher level of autonomy and independence from state intervention.

2.2. Main features of the rail restructuring policies

The worldwide restructuring process of the rail industry began with timid reforms. Many countries began by replacing their national railways with autonomous commercial bodies possessing independent, realistic balance-sheets, in which only public service obligations could be explicitly subsidized by the government. Other countries opted to substitute their old geographically-based management with a multi-divisional structure, defined by the companies' different lines of business or services.

A common feature in most cases is that some countries have carried out a relatively long-term restructuring, whereas others have preferred a quicker implementation. For example, privatization in New Zealand and Japan was phased over several years, while Argentina and the United Kingdom took less than two years. Another common characteristic is that all restructuring processes were undertaken to make the companies attractive to private investors, although full privatization has been less preferred than concessioning.

The changes have involved the revision of laws and other regulations affecting railways, reducing staff, dealing with pension issues, and deciding how much property should be sold and how much should be retained by the state. In addition, several arrangements for paying for unprofitable (but socially needed) train services were put into place, together with a precise definition of the concession contracts and their main terms.

With regard to results, in general, most of the restructuring experiences detailed below seem to have been positive. The objectives of stopping the industry's drain on the state's resources, along with the stabilization of market share for both passengers and freight, were achieved in most countries. Likewise, the companies succeeded in raising their levels of productivity.

3. New organizational models for the rail industry

Despite all these changes, the most salient characteristic of the restructuring process of the rail industry in the last decades has been the consolidation of different and alternative organizational structures for the industry as a whole. These structures differ along three main features to be analyzed in detail: how are access and infrastructure and multimodal competition considered, what is the extent of vertical separation introduced after the change, and what is the amount of private participation allowed in the industry after the reform.

3.1. Access to rail infrastructure and intermodal competition

The management of rail infrastructure not only includes simple pricing principles, but also access rights and long-term development provisions. Each country addresses these differently: most have opted to publicly retain infrastructure, creating state management agencies (Sweden's *Banverket*) to regulate private train operators (as in Argentina); others (France, Germany or Spain) have established nominally independent but state-owned enterprises to manage stations and tracks. Only the United Kingdom privatized

infrastructure and operations in 1996, although part of the changes were later reverted in 2001. The financial collapse of *Railtrack* (the private owner of the infrastructure) and the poor infrastructure maintenance (that provoked serious accidents and significant disruptions to service) were the main reasons to dissolve *Railtrack* and substitute it for a *not for profit* body, *Network Rail*. This suggests that, whether in public or private hands, rail infrastructure regulation must include minimum investment requirements to avoid short-term myopia and to ensure that key investments are prioritized over dividend increases or defence against potential takeover.

On the other hand, the separation of infrastructure from services also implies that the new models should focus on the issue of access, which is particularly relevant in the case of highly integrated trans-national networks (as in Europe) or privately or publicly managed dense networks (as in the United States, Canada and some Asian countries). In the European Union, for example, Directive 91/440 directs each member state to grant international access and transit rights to international groups where stakes are held by railway undertakings in that or other member states. There have been no directives or resolutions related to domestic traffic, although the European Commission advocates the extension of these provisions to all freight and international passenger services. In January 2002 the Commission adopted a new communication: *towards an integrated European railway area* (known as the *second railway package*). Open access to the infrastructure for national services is promoted in order to completely open up the rail freight market. It has been agreed and open access in the domestic freight market will be introduced in 2007. However, open access in the passenger market is a much slower process.

In the privatized structure of the United Kingdom, open access to passenger services has been limited by a number of provisions regarding that moderate competition. Initially designed to protect rail franchisees from new entrants and from each other, these provisions were anticipated to be gradually reduced over time. In other countries (Argentina and Côte d'Ivoire-Burkina Faso), access rights are also clearly specified in the contract. In certain large cities, like Mexico, D.F. or Buenos Aires, operators share a common network under a unique transport authority.

The final aspect regarding access rights to rail infrastructure lies in the removal of existing or potential barriers to entry that might distort competition by favouring some competitors over others. These barriers also include technical requirements (for example, those related to incompatible rolling stock and tracks) and safety standards (in terms of a common minimum level). In summary, the general rule should be to promote open access as widely as possible once the separation between the natural monopoly infrastructure and train operations has been effectively achieved. However, this process must depend upon a detailed analysis of infrastructure costs and the prices charged to cover them.

Barriers to entry are also related with intermodal competition. As already mentioned, modal choices can be heavily distorted due to different cost coverage ratios and the use of different cost input bases. A solution is to follow an integrated, multi-modal approach. Basic principles will have to apply to all transport operators, irrespective of the mode in which they operate. For example, in countries like Argentina and Chile, the extent of road freight transport competition was considered in designing the rail concession contracts. The general rule was that operators undertaking business at their own commercial and financial risk should not be at an undue disadvantage to those who enjoy public aid or indirectly benefit from huge externalities.

In the case of rail infrastructures, the principles envisaged to avoid these distortive effects should be solidified in the coordination of existing networks (particularly in dense rail areas) and the establishment of mechanisms that facilitate inter-operability and international links. However, not even the most advanced infrastructure regulations, such as the Swedish and the British systems, offer much help since they were conceived for a single-country environment. In other countries, such as Argentina before the restructuring process, railways attempted to solve national transport problems (by offering under-priced passenger services or subsidized low-quality freight transport). As a result, their financial performance rapidly deteriorated in an isolated framework. Therefore, the infrastructure pricing strategy in these areas should be compatible with the achievement of both local and international objectives, by establishing, if needed, a system of slot assignments in more congested corridors.

3.2. The degree of vertical separation

One of the most clearly defined patterns emerging from deregulation and restructuring is that they carry out two critical dimensions summarized in Table 2: the degree of *vertical separation* between infrastructure and services, and the involvement of *private management* in the sector. With respect to the first dimension, there are three main options for the vertical organization of the railway industry: (i) *vertical integration*, (ii) *competitive access*, and (iii) *vertical separation*.

The first option corresponds to the traditional, historic model of railway organization described above, where a single (usually public) entity controls all the infrastructure facilities as well as the operating and administrative functions. Less frequent, competitive access is characterized by the existence of an integrated operator, who is required to make rail facilities (tracks, stations, etc.) available to other operators on a fair and equal basis through the trading of, for example, circulation rights. This has the advantages of integration (economies of scope, coordinated planning and reduction of transaction costs), but its overall effectiveness may be jeopardized if the integrated company has incentives to leave out other operators.

Alternatively, in the complete vertical separation scenario, the management (and, possibly, the ownership) of facilities is fully separated from other rail functions. This is very attractive because although infrastructure may remain a natural monopoly, it is separated from rail services, where potential competition among different operators is possible. In general, the main advantage of this vertical unbundling is that rail transport is placed in a similar situation as road transport, especially regarding the tariff system and infrastructure planning. Governments could study investment proposals on the basis of a cost-benefit analysis, while pricing policies could be based on social cost.⁴ In addition, separating infrastructure from services greatly facilitates the entry of more than one operator on a single route. For profitable services this would permit notable

⁴ Note that an important problem here is the difficulty of defining the social cost of use of railway infrastructure. Determining the marginal or incremental costs of the use and wear and tear of one additional train is not, in principle, any more difficult than the equivalent calculation for road transport. The problem, however, is greatly complicated for the railway when this cost is evaluated in a congested environment. In pure economic terms, this cost is the opportunity cost of the stretch of track in question, but in practice, it is difficult to quantify this opportunity cost, especially if there is a mixture of social and commercial services.

improvements in efficiency by allowing direct competition among operators. For non-profitable services, infrastructure separation can be accompanied by tendering, to stimulate increased efficiency through competition for the market, the introduction of innovations, and marketing improvements.

Table 2: Alternative organizational structures for the rail industry.

		<i>VERTICAL UNBUNDLING</i>		
		Total vertical integration	Competitive access	Vertical separation
PRIVATE PARTICIPATION	Government Department	India, China, former socialist countries.		
	Public Enterprise	European railways		
	Reformed Public Enterprise	Many European railways at present		Sweden
	Service Contract with Private Sector		Japan (HSR) US (rolling stock) Pakistan (ticket sales)	UK (rolling stock)
	Management Contract with Private Sector	Nigeria (1980)	US small railways	
	Leasing to Private Sector		Amtrak (USA) (track) VIA (Canada) (track) Japan (track) Cameroon (baggage)	
	Leasing from Private Sector		US and Europe (wagons and cars)	
	Concession (franchising)		Argentina, Brazil, Chile, Cote d'Ivoire	UK (passengers)
	Joint Venture		Canada US (pipe and wire)	UK
	Private Company	New Zealand	Japan, US (Class I), Canada	UK (freight, infrastructure)

Source: Elaborated from Galenson and Thompson (1993).

However, the vertical unbundling of the rail industry also implies several disadvantages. The main problem is the potential loss of economies of scope derived from the joint operation of tracks and services. It is often noted that the relationship between the services supplied and the rolling stock used, as well as the quality, quantity and technical characteristics of the infrastructure, is so close that both aspects need to be planned together. Thus, assigning different services to several operators may decrease the utilization of the sector's staff and physical assets. Another negative factor is the higher risk that the new system becomes less attractive to the user than an integrated system.⁵ It is also mentioned that vertical separation requires such a complex institutional arrangement that the resulting transaction costs will be often prohibitive for many countries. A final disadvantage of vertical separation is the reduction of investment incentives. For example, an infrastructure owner considering an investment on a facility with only one potential buyer will anticipate bargaining away some of the benefit from the new service once it comes on line. This problem becomes less relevant with more competition in the market, since competition weakens the bargaining position of individual operators by reducing the specificity of the assets.

⁵ For example, because of the lack of interchangeable ticketing, an integrated national network, etc.

3.3. *The amount of private participation*

With respect to the dimension of private participation in the industry, Galenson and Thompson (1993) provide a list (ordered in terms of increasing private participation) of the different situations that can be found in the world's rail industry. The first situation is a *government department*, where the railroad is fully controlled and financed by the government and therefore subordinated to its interests.

The second example is a *public enterprise*, where the railway is characterized by a higher managerial autonomy, but is still requires government approval for many decisions. Normally, these railways sign *contracts* (or have sectoral laws) with the government, specifying each party's objectives and attributions and the financing rules. Similarly, the case of a *reformed public enterprise* corresponds to a situation where the railway is incorporated (into a shareholding company), commercialized (financially and managerially autonomous), and made subject to the country's company law. However, the government, as the main owner, determines pricing policies and investment levels, while guaranteeing the supply of non-economical social services with the necessary subsidies.

There are other situations that include mixed forms of cooperation between private and public capital. For example, rail service in some countries is provided through a *service contract* with the private sector, where, maintaining full ownership, governments or public enterprises can contract activities to be performed by private sector entities, including food catering, medical services, ticket sales, maintenance of physical assets, etc. Related to these there are *management contracts* with the private sector, where the contractor assumes responsibility for the operations and maintenance of certain activities. One variation is *leasing* to the private sector, where the contractor pays a fee for the use of the fixed assets. The lease contractor has more autonomy than in management contracts, controlling aspects such as the working capital and staff, but also assumes more risk. The owner maintains responsibility for investment and debt service. In many countries, locomotives and wagons are sold or leased to non-railway entities for transporting very specialized goods.

Concessions are a broader form of lease where the contractor also agrees to make certain fixed investments and maintains the use of the assets for a longer period. This is currently the preferred restructuring method in the rail industry and will be extensively discussed in the rest of this chapter. Finally, *joint ventures* entail the largest degree of private participation. Private partners contribute development capital and planning and management expertise to develop land or other real estate owned by a railway. Also, under full *private ownership*, certain services or whole companies are operated by private firms.

4. **New regulatory scenarios for the rail industry**

The vertical separation/private participation bi-dimensional space discussed above creates a new regulatory framework in the rail sector whose most relevant characteristic is the flexibility. It introduces significant new roles and functions for the regulator and modifies the number of possible regulatory structures and models. In practice, choosing a particular method for railway restructuring depends on a number of particular

objectives or goals that the Government must balance according to the economic environment in which it operates.

One of the first elements to consider is the existence of *financial constraints*. If they are important, the maximization of the proceeds obtained from the restructuring process will be a primary goal. A second element to consider is the pursuit of *internal (or cost) efficiency* in terms of providing services at the lowest possible cost, and therefore generating an efficient use of resources. Similarly, there is the goal of attaining *allocative efficiency* by setting optimal prices equal to the marginal social cost, which from an intermodal viewpoint, facilitates the best distribution of traffic. The objective of *dynamic efficiency* requires the long run minimization of cost through active, technology-improving investment policies. There can also be *equity objectives*, such as facilitating transport for all citizens, independent of income level. Finally, the government can also consider the optimal *allocation of capacity*, which favours management of railway capacity, coordination with other modes of transport, and overall *minimization of risks* in terms of service maintenance over time, risk of default, etc.

Table 3: Different rail regulatory scenarios and their objectives.

Scenario	<i>Objectives</i>						
	Financial constraints	Internal Efficiency	External Efficiency	Dynamic Efficiency	Risk Minimizing	Capacity Allocation	Equity
(1) Vertical Integration and Government Department	-	-	+	+	+	+	+
(2) Vertical Integration and reformed public enterprise	-	-	+	+	+	+	+
(3) Vertical Separation and Reformed public enterprise	-	-	+	+	+	-	+
(4) Competitive Access and Concession regime	+	+	+	unclear	unclear	-	+
(5) Vertical separation and Concession regime	+	+	+			-	+
(6) Vertical integration and Private enterprise	+	+	-		-	+	-
(7) Competitive access and Private enterprise	+	+	-		-	-	-
(8) Vertical Separation and Private company	+	+	-	-	-	-	

Note. A “+” sign means that the objective is easily achievable within the corresponding scenario. The “-” sign implies the opposite.

Table 3 summarizes the combination of these objectives, creating at least eight different possible regulatory scenarios, grouped in decreasing order of private

participation. Some additional scenarios, such as the mixed forms described above, have not been included.⁶

It is important to note that the objectives could be given a different weight. For example, financial and cost efficiency objectives are now valued above all others, which explains the privatization boom, through concessions and direct sales to the private sector. In addition, as the degree of privatization increases, there is a trade-off between social and financial efficiency objectives. The public company scenarios serve social objectives (equity, reduction of risk on the service, intermodal coordination, etc.), but there are inefficient, leading to huge commercial deficits. As we have already indicated, this was the main reason for the restructuring of the sector.

The deregulation measures that define scenarios 4 and 5 (*concessions*) have the advantage of favouring the efficiency and solvency of the companies, as well as reducing the state's financial burden (although these effects are possibly not as great as with direct privatization). In addition, concession contracts allow the cushioning of some of the negative effects that may arise from the actions of the private company. Thus, it is habitual to establish maximum prices and minimum service levels so that impact on equity can be minimized. Likewise, many routes which, though not profitable, are beneficial from a social viewpoint can continue to be served: concessioning them to operators who request lower public subsidies meets both efficiency and equity objectives.

In regard to dynamic efficiency, the first results of the investments implemented by the restructured companies or bodies are ambiguous. In Argentina, the investment levels of some operators have been below those foreseen in their concession contracts, though at the aggregate level, investment levels seem to have improved. Something similar has occurred with some passenger franchises in the United Kingdom. At any rate, the effective investment levels should be compared with those that existed in the regulated context. In this sense, other experiences have indeed led to a substantial recovery in investments in both infrastructure and rolling stock, as well as an improvement in service quality. In other countries, such as Japan, privatization does not seem to have slowed the technological development of the railway industry (Fujimori, 1997).

Apart from other considerations, operational risks are minimized when entrusted to a public enterprise. With a private company, there is obviously a greater risk of closure of certain services, or of larger instability. Again, concession systems allow the risks inherent to the action of private enterprise to be reduced.

Finally, the problem associated with managing capacity is easily eliminated in the case of vertically integrated companies, although this is not so simple for systems of competitive access or separation. In this case, the problem is increased for companies with high traffic densities and conflicting capacity demands. Modern computer technology can reduce the problem through real-time management of electronic systems, but when connecting systems have different informational qualities and dispatching priorities it is very difficult for anyone to plan and manage integrated services across several systems.

⁶ This is because many of these forms of private participation are related to very specific services (e.g., the case of *service* or *management contracts*) and on occasions some of the forms of contracting (e.g., *leasing*) are very similar to those established in a concession or franchising system.

5. The role of concession contracts in the rail industry

In spite of the number of potential regulatory scenarios just described, few railways around the world have been fully privatized. Instead, most countries have opted to concession rail services and even rail infrastructures in some cases, to private firms in exchange for a fixed payment. This has been the favoured form of restructuring because it allows the government to retain ultimate control over the assets while the private sector carries out day-to-day operations according to pre-specified rules devised in a contract that transforms the problems associated with traditional regulation into issues of contract enforcement.⁷

Since there are many variables to consider, rail concession contracts cannot be reduced into a single standard model. However, according to existing experiences, Table 4 proposes four key variables to consider.

Table 4: Some key variables in rail concession contracts design.

Type of contract	<p><i>Package size depends on economies of scale/scope and existing potential for competition</i></p> <p><i>Horizontal concessions (geographic) according to country's characteristics</i></p> <p><i>Vertical concessions (functional) according to network's characteristics (including current state of infrastructure and new investment needed)</i></p> <p><i>Mixed packages depending on profitability and bidders' financial constraints</i></p> <p><i>Freight vs. passenger concessions depending on relative traffic shares</i></p>
Award and duration	<p>Pre-qualification requirements to reduce risks</p> <p>Type of auction (sealed, one-shot) and explicit rules for auctioning</p> <p>Selection based on government's objectives (fiscal, equity or efficiency)</p> <p>Short periods (favour competition; diminish investment incentives) versus long periods (favour investment; diminish enforceability)</p> <p>Termination: re-auction preferable to automatic renewal</p>
Operating general contents	<p>Concessionaire:</p> <ul style="list-style-type: none"> obligations: services (with adequate performance) and payments rights: exclusivity and compensation for public service obligations <p>Government:</p> <ul style="list-style-type: none"> risk sharing (net cost/gross cost mechanisms) asset ownership rules
Regulation mechanisms	<p>Price control rules (services and infrastructure)</p> <p>Principles regarding price discrimination and cross-subsidization</p> <p>Definition of quality targets and quality control</p> <p>Issues regarding safety and externalities</p>

The first critical aspect of a concession is determining its *type*, both in vertical (functional) size and horizontal (geographical) size. Recent concessions in the rail

⁷ The list of countries with actual or planned rail concessions include, among others, United Kingdom, Argentina, Chile, Brazil, Bolivia, Peru, Colombia, Guatemala, Mexico, Côte d'Ivoire-Burkina Faso, Cameroon, Congo, Malawi, Jordan and Mozambique.

industry have created smaller horizontal packages throughout the country. For example, rail freight systems in Argentina, Brazil, Mexico and Colombia were split into several regional companies, and Chilean railways were broken down into four passenger companies and two freight companies with a separate infrastructure firm. All of these countries also used economic criteria to design the size of the concessioning package, accounting for the profitability of different lines.

In Europe, functional separation between infrastructure and services has been preferred, especially after European Commission Directive 91/440. At its most extreme, this form of concessioning was used in the privatization of British Rail, which also included the private provision and management of rail infrastructures. A less extensive vertical separation has been developed in Sweden and other European countries, where infrastructure has not been auctioned off to private firms (Lundberg, 1996). However the debate about the advantages and disadvantages of the separation of infrastructure and operations is not closed. There is a perception that separation is an essential condition for non discriminatory access, and that it is very difficult to increase competition in a situation in which the major operator controls the infrastructure. But the problems of British process have increased the doubts about its advisability. Nash (2004) points that perhaps the Swedish model, combining a stated-owned infrastructure entity and a charging based on short run marginal cost, is the system that has worked with best results in Europe.

The second key issue in designing rail service and infrastructure concession contracts is defining the *award process* and *duration* of the concession. This includes the auction rules and, particularly, the criteria defining how each concession will be awarded to a private operator. There are a number of possibilities to choose from as the award criteria (for example, maximum payment to government or minimum tariff). There is also a choice between unrestricted bidding and bidding that could involve some pre-selection (see Guislain and Kerf, 1995, and Kerf *et al.*, 1997). In the privatization of the former British Rail, for example, the concession process began with a pre-qualification stage, followed by a formal invitation to tender for a particular package. After indicative bids were received, four bidders were short-listed. One of these was subsequently named the preferred bidder, and was given a fortnight to complete financing and other organizational arrangements before being confirmed as the winner. At that point, the regulator gave public details of the bid, in terms of the required subsidy and promised service improvements.

With respect to bidding mechanisms, there is extensive literature on experiences and results in different auction forms. Single, sealed-envelope bids is the simplest, avoiding collusion and obtaining higher bids. However, more complex approaches, such as real-time auctions, have been used in some transport concessions. Once the rules have been set up and the bids requested, bidders should have a study period to form their own evaluation of the potential gains to be extracted from the concession. Early research by Preston *et al.* (1996) for the United Kingdom indicated that key issues for bidders were the length of franchises, the level of competition they would face from other operators, the separation of infrastructure from services, the costs (including new investments) associated with maintenance and the selection criteria for the bidding process.

Although the guiding principle should be to maximize competition so that the most efficient firm ends up winning the award, it is clear that there is no single method for selecting the winner once bids have been submitted. The final choice depends on the Government's objectives, which should be explicit and built on transparent criteria.

Thus, if the government intends private participation to be a means of reducing the burden on the public sector, it must use fiscal benefits as the main criterion, looking at who requires the lowest subsidy or who offers the highest auction price. In Brazil, for example, the six regional rail concessions were successfully auctioned to the highest bid above the Government's minimum price. Concessionaires were required to make an up-front payment immediately after the auction, followed by a stream of pre-determined payments over the life of the concession. Similarly, in Britain, minimizing subsidy payments appeared to drive the regulator's choice of bidders, especially in the first concessions. Other criteria were the financial position of the tenderer, its managerial competence and its operational proposals.

Alternatively, if tariffs and quality of service are defined in the contract, bids can be evaluated on the basis of the lower cost provider, simultaneously including penalties for not achieving certain performance objectives. Social objectives can be also targeted by focusing on the bids that propose to monopolize the industry for the lowest number of years or to charge the lowest fare to final users. Sometimes, as in the case of rail freight, the traffic mix makes the price structure very complex, so that this mechanism becomes impractical. Moreover, using tariffs as an award criterion for rail concessions limits the later possibility of regulatory intervention in prices and demands an adequate definition of quality standards.

Many concessions in the rail industry have been awarded using formulas with multiple criteria, which can account for a larger number of objectives. For example in Argentina, the bids for the six freight packages that were concessioned were evaluated using the net present value of the canon to be paid to the government during the first fifteen years of the concession, the quality of business and investment plans, staffing levels, the proposed track fee for passenger trains, and the share of Argentine interest in the consortium. The weights of these criteria reflected both the importance attributed to investment in the railways and political compromises on employment. However, for the award of metropolitan commuter railways, the Argentinean authorities kept things simpler to make the bidding process and final selection as transparent as possible. They learned from the freight concession that selecting the winning bid through numerous cumbersome criteria with discretionary weights was more likely to reduce the efficiency of the bidding process than to improve it. Instead, the terms of the concession should be made clear to all potential bidders and bidding should take place on the basis of a single parameter encompassed in the bidders' economic assumptions in terms of the concession.⁸

With regard to the optimal *duration* of the concession contract, the trade-off is evident in terms of efficiency, since the shorter the concession, the more immediate the competitive pressure, but the lower the incentive to invest and develop the business. Longer concessions, in contrast, tend to diminish the regulator's enforcement capacity and soften the incentives to promote efficient outcomes. The general rule is to adapt the concession period to the economic life of the assets and to make this compatible with the government's objectives. This balance often creates conflict: while concessionaires generally argue for long contracts that provide them with incentives to build up the business and purchase or replace long-lived assets, concessioning authorities prefer shorter lengths to favour the achievement of efficiency (by the implicit threat of not

⁸ In the case of the metropolitan railway concession, for instance, each concessionaire calculated her expected revenue from operations, then compared it with the capital investment programs and finally estimated the subsidy amount to be requested (The World Bank, 1996).

renewal) and fiscal goals (since the canon or auction price may be increased after the first few years of the concession). Only if sunk investments are minimal and asset reutilization is possible, are shorter periods advisable for particular rail services (those related to signals, track and station maintenance).

Shaw, Gwilliam and Thompson (1996) point out that the average duration of a rail service concession is five to ten years, increasing up to thirty when network investment and development are included. In Argentina, for example, the six freight packages were concessioned on a thirty year term, with an optional ten year extension, due to the poor state of infrastructure and the huge investment that was required. For similar reasons, the international rail link between Côte d'Ivoire and Burkina Faso was awarded in a fifteen year concession. Conversely, train operating companies in the United Kingdom were granted a concession to run passenger services for a period of only seven to fifteen years.

After the duration period has expired, the contract must also specify several *termination arrangements* to avoid any disruption in services. One possibility is to make automatic renewals in case new candidates for the concession do not exist. The regulator should not compromise on this before the concession ends in order to ensure that the incumbent has the correct incentives. New auctioning seems to be the standard procedure after a concession has ended, but most rail operators will seek a renegotiation of duration terms while the contract is still in force. Examples of this strategy are some United Kingdom rail franchises who argued that they made long-lived investments in high-quality wagons and locomotives when they asked for a license extension.

Since renegotiation costs money, but a lack of renegotiation might cause performance deterioration, concession contracts should specify the circumstances for renegotiation, and which party should initiate the process. If intermediate objectives are achieved, a pre-scheduled revision process might help to reduce both parties' risks. Although the contract will always be incomplete, standard clauses should include behaviour in unforeseen changes in demand conditions, responses to unanticipated rises in energy or labour costs, etc. For example in Argentina, freight concessions could not fulfil their promise to invest \$1.2 billion in the rail network over fifteen years due to unexpected falling traffic levels.

A flexible contract renegotiation mechanism is a good idea in any case since the Government may face the dilemma of enforcing contracts to the detriment of the operating companies and the national rail system or rescheduling investment and making other compromises at the cost of undermining his credibility for enforcing future agreements (Carbajo and Estache, 1996).

This is why one of the most critical issues in designing a rail concession contract is the *attribution of rights and obligations* to the parties. On one side, the private operator pays a regular canon or receives a subsidy and is awarded the right to operate train services and/or manage its infrastructure (including future investments) with (total or partial) exclusivity rights that protect her from other competitors. On the other hand, in exchange for the payment or the compensating subsidy, there is a regulatory activity by means of which the overall performance of the sector is monitored and a stable framework for current and future rail operations is provided.

These operations may include infrastructure provisions if they were auctioned off to private firms. In fact, a large part of railway activities might be concessioned. These include infrastructure: track, signals, stations, yards and shops; operating equipment: locomotives, wagons, carriages; and general service access to track, route and schedule

information and maintenance. The exact form in which this process is developed in practice depends on the parties' risk sharing agreements. According to a service contract, for example, train operators provide rail transport services for passengers or (rarely) freight according to specific routes, levels of quality and technology as established by the regulator. The operators may cover some investment costs and carry some commercial risk, which can be integrated into a net cost contract, where the operator keeps all revenues generated by passenger or freight traffic. This type of contract, where the operator carries revenue as well as cost risk, often generates more traffic and is let to the most attractive bid, but offers a higher incentive to predate. Alternatively, gross cost contracts specify that all revenue accrues to the government and the contracts are let on the basis of the least total cost supplier so operators carry cost but not revenue risk. The experience in the United Kingdom with regard to passenger franchises suggests that gross cost contracts generate more bids per tender (particularly from new entrants), offer greater incentives to public revenue generation, reduce the administrative cost for the regulatory authority, and support any fare scheme with modal integration and quality control.

The regulator may retain control over and responsibility for common functions, and its main roles should be restricted to regulating quality (in terms of service, safety, environmental and technical standards), controlling monopolistic behaviour (in terms of abusive prices or services), and determining the overall characteristics of the function of the sector (in terms of coordination at the national and international level) according to established the competition rules or rights and anti-trust and commercial legislation.

The implementation of rail concession exclusivity rights varies in each country. In Argentina, freight concessionaires have exclusive use of tracks but must grant access to passenger operations in return for a compensatory track fee. In Chile, passenger services and infrastructure initially remained in public hands, while freight concessions were awarded to private competing firms. The fifteen-year concession for the Côte d'Ivoire-Burkina Faso trans-national railway was awarded with a seven year exclusivity period, after which the operator should grant access to third-parties specified by the regulator for an agreed fee. Thus, exclusivity rights should be viewed as another instrument for regulatory control, and not taken for granted by the firms *ex-ante*. Limiting the duration of the monopoly period balances the regulator's desire to reap the benefits of competitive access to the tracks and the private train operators' preference for full control of the market to generate profit and facilitate revenue forecasting. In general, most railways have been concessioned on an exclusive basis in geographical areas, as in Argentina or Brazil, possibly with some access rights for connecting railways to certain central or strategic track segments. This has been due to the geopolitical configuration of the country, the density of the existing network, and the need to promote competition in major markets (as in Mexico) or for non-competing services (such as passenger services on freight tracks in Chile).

With respect to the concessionaires' obligations, the private provision of rail transport services, particularly in less developed areas or zones with a structural lack of network, cannot always be separated from public subsidization or reciprocal compensation for politically motivated public service obligations. Arrangements for these loss-making but socially necessary services must be included in concession contracts, in terms of detailed performance levels to be attained by the firm, possibly even be designed to be awarded to the company willing to provide the specified services for the lowest level of subsidy (negative concessions), as in Argentina.

A final feature of defining the rights and obligations of the concessionaires, the current experience of rail concessions in South America shows that restructuring has often lowered employment levels. This is, in practice, one of the toughest obstacles hindering the private participation process in certain countries and often requires difficult political decisions. In Brazil, for example, large redundancies were inevitable and were dealt with in two phases. Before concessioning, incentive schemes for early retirements were in place; after the concession was awarded, the former national rail operator paid involuntary separation grants to the remaining staff not hired by the concessionaire. After that point, compensation for additional laid off employees is the responsibility of the private operator. Undoubtedly, any such employment constraints will be reflected in the auction price of the concession.

In summary, in their general form, rail concessions are the most advantageous solution to the challenges posed by the current regulatory environment of the rail industry. It usually adopts the form of a long or medium term contract where a vertically or horizontally integrated package of (passenger and/or freight) rail services is auctioned off to private firms, while economic assets remain public property. Three of its key features – type, duration and contents – have been described in this section, but there are other particular aspects of the concession contract design in the rail industry that, based on their importance, deserve a more detailed treatment. These include price regulation, in terms of defining the most important issues in establishing effective and well-oriented price control mechanisms; quality regulation, in both its static dimension (quality of service, safety and environmental issues) and dynamic dimension (rules for infrastructure investment and financing), and coordination between infrastructure and superstructure.

6. Conclusions

The increasing role of private sector in the rail industry is one of the most relevant characteristics of the evolution of this industry in recent years. This change is reshaping the way in which Governments are addressing the main regulatory challenges derived from the economic and technical characteristics of railways.

In this paper we have showed that the industry regulation is moving accordingly towards more flexible schemes of public intervention. Although this does not fully preclude direct participation by the Government, it seems that the monopolistic rail company is progressively disappearing as the dominant model around the world. There is no unique form of rail regulation to address these new challenges, but the general rule is to maintain flexibility and simplicity whenever possible.

Two key issues in the new regulatory environment of the rail industry are that private participation is included in license contracts and the organization of the industry is adapted to each country's needs and characteristics. In turn, the use of these mechanisms also changes the role of the rail regulator, whose actions should now be governed by principles that foster competition and market mechanisms and simultaneously provide a stable legal and institutional framework for economic activity. The regulator should refrain from intervention unless the ultimate goal of achieving economic efficiency subject to the socially demanded level of equity is in jeopardy.

Nevertheless, two important caveats for future regulation must be taken into account. First, the process of privatization chosen in each country depends on the basic objectives sought: to maintain an industry with one operator or a small number of operators, or to facilitate a process of competition on the track. Second, legacies from the traditional mechanisms of regulation should be avoided. In particular, high debt levels and overstaffing are two common problems that must be dealt with before starting any privatization policy.

In any case, future researches will be necessary to evaluate the advantages and difficulties of the current rail restructuring processes in the world. Some of these costs and benefits have been described in the paper. The new regulatory schemes will be essential in order to preserve the advantages of these new rail systems and to reduce its potential problems and costs.

References

- Baumol, W. J. (1977). "On the Proper Cost Tests for Natural Monopoly in a Multiproduct Industry". *American Economic Review*, 65, 5, pp.810-822.
- Baumol, W. J., Panzar, J. C. and Willig R. D. (1982) *Contestable Markets and the Theory of Industry Structure*, Harcourt Brace: Jovanovich, New York.
- Carbajo, J. C. and Estache A. (1996) "Railway Concessions – Heading down the Right Track in Argentina", *Viewpoint* 88. Finance and Private Sector Development Department. The World Bank: Washington, D.C.
- Fujimori, Y. (1997) "Ten Years of JNR's Privatization (Technological Development after JNR's Reformation)", *Japanese Railway Engineering*, 138, pp.3-5.
- Galenson, A. and Thompson, L.S. (1993) "Forms of Private Sector Participation in Railways", *TWU Papers* No. 9. The World Bank: Washington, D.C.
- Gathon, H. J. and Pestieau, P. (1995) "Decomposing Efficiency into its Managerial and its Regulatory Components: The Case of European Railways", *European Journal of Operational Research*, 12, pp.500-507.
- Guislain, P. and Kerf, M. (1995) "Concessions – The Way to Privatize Infrastructure Sector Monopolies", *Viewpoint* 59. Finance and Private Sector Development Department. The World Bank: Washington, D.C.
- Kopicki, R. and Thompson, L. S. (1995) *Best Methods of Railway Restructuring and Privatization*, CFS Discussion Paper Series No. 111. The World Bank: Washington, D.C.
- Lundberg, A. (1996) "Restructuring of the Swedish State Railways", *Japan Railways and Transport Review*, 218. pp.22-26.
- Nash, C. A. and Rivera-Trujillo, C. (2004) "Rail regulatory reform in Europe- principles and practice", STELLA Focus Group 5 Synthesis Meeting, Athens.
- Oum, T. H. and Yu, C. (1994) "A Comparative Study of OECD Countries' Railways", *Journal of Transport Economics and Policy*, 38, pp.121-138.
- Preston, J., Whelan, G., Nash, C. A. and Wardman, M. (1996) *The Franchising of Passenger Rail Services in Britain*, Institute of Transport Studies: University of Leeds.
- Shaw N. L., Gwilliam, K. M. and Thompson, L. S. (1996) "Concessions in Transport", *TWU Papers*, No. 27. The World Bank: Washington, D.C.
- Thompson, L. S. and Budin, K. J. (1997) "Global Trend to Railway Concessions Delivering Positive Results", *Public Policy for the Private Sector*, No. 134. The World Bank; Washington, D.C.
- Waters II, W. G. (1985) "Rail Cost Analysis". In Button, K. J. and Pittfield, D. E. (eds.). *International Railway Economics*. Gower: Aldershot.

Acknowledgements

Pedro Cantos thanks financial support from Spanish Ministry of Science and Technology under project SEJ 2004-00110. Javier Campos gratefully acknowledges financial support from the Spanish Ministry of Science and Technology and from FEDER through grant BEC2002-02527. Both authors are grateful to The World Bank for collaborating in previous versions of this paper.