Rail Transport Regulation*

Javier Campos
University of Las Palmas
(Spain)

Pedro Cantos
University of Valencia
(Spain)

* This paper has been prepared for the Economic Development Institute of the World Bank. The authors are grateful to Ginés de Rus, Antonio Estache, Andrés Gómez-Lobo and Lou Thompson for valuable suggestions and comments. They also acknowledge useful discussions with other colleagues at the University of Las Palmas and the University of Valencia.

LIST OF BOXES

1. – THE ECONOMIC CHARACTERISTICS OF RAIL TRANSPORT.

The railway 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 played by infrastructures and networks, the existence of indivisibilities in inputs and outputs, the organization of the rail transport as a public service, and the existence of externalities in the transport system as a whole (Button, 1993). These characteristics not only define a descriptive framework for this sector that will be continually referred to throughout this paper, but also jointly determine the main factors that should be considered when studying in detail the appropriate economic regulation for the rail industry.

1.1. – The multi-product nature of the 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, together 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 coexists with local services (suburban and commuter trains), regional services, and even with high-speed trains on certain corridors or routes.[1]

The implications of the multi-product nature of the activity can be examined at different levels. At the accounting level, for example, it is often difficult to allocate total operating costs among services. For instance, 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 coexist with other costs not affected by changes in output.[2] Some cost elements may be attributable to a particular traffic (for example, passengers), but most of them (wagons, energy, staff,...) may 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 be considered in the multi-product setup of the rail industry is the sub-additivity of the cost function faced by a railroad. 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. 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 is connected with the advantages and disadvantages of the separation of infrastructure from services and will be discussed below in depth after first describing in more detail the cost structure of a typical

[1] Commuter and suburban passenger traffic will be analyzed in other working-papers on this series.
[2] 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.

railway company.

1.2. – The cost structure of railways companies.

Railways costs are often classified into four broad cost categories: *(i)* train working costs, which include the costs of the provision of transport services (fuel, crew, maintenance and depreciation of rolling stock); *(ii)* track and signaling costs (including the operation, maintenance and depreciation costs of the infrastructure); *(iii)* terminal and station costs; and finally *(iv)*, administration costs (Waters, 1985). The first two cost categories, typically, are prevalent in most companies and change according to several factors.[3]

Among train working costs, the cost of rolling stock items depend on both their amount and the distance they run. Fuel costs depend on car-kilometers run for each type of vehicle, while train crew costs vary according to train-kilometers run. Track and signaling costs usually rely on the length of the route (since they all usually rely on a single, standard-quality track). The amount of track and signaling required, however, changes with the number of trains for which paths are required, although this relationship is not constant. Terminal and station costs depend on the traffic volume, but they vary considerably with the type of traffic.[4] Finally, administration costs fluctuate depending on the overall size of the firm, although the precise nature of this dependence is difficult to determine in general.

Therefore, as mentioned above, allocating the different costs of a typical rail operator to the multiple outputs or inputs it produces is a complex task. It often involves a degree of arbitrariness that requires, from a regulatory point of view, a clear distinction between costs that are avoidable and those that are not. The avoidable costs are uniquely associated with a particular output: were this output not produced, no cost would be incurred. This guiding principle relates to the cost recovery for particular outputs. Avoidable costs may therefore be considered as a floor to regulated prices (if any), since charging less than the avoidable cost would be equivalent to operating at an economic loss.

1.3. – The particular 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 was its efficiency. The existence of substantial fixed costs (particularly, those associated with infrastructures) traditionally led economists to assume the presence of important economies of scale in this industry, and thus the provision of rail transport services was typically regarded as a textbook example of a natural monopoly.

However, this notion has been heavily challenged in recent decades and a number of new ideas have been brought to the economic analysis of this industry. The upheaval and development of the theory of contestable markets (Baumol, Panzar and Willig, 1982)

[3] Nash (1982) finds that train working costs in European firms (with the notable exception of high speed passenger traffic) accounted for 44%-45%, whereas track and signaling was just 23%-26% of total costs.

[4] For instance, bulk freight handling requires more terminal expenses than parcel services. Similarly, long distance passengers require more services (ticketage, reservations, luggage,...) than short distance users.

contributed to clarifying the proper definition of the natural monopoly concept, in terms of the cost function being sub-additive. This concept implies that whereas duplicating rail infrastructure is generally inefficient (therefore categorizing the rail network as subject to natural monopoly conditions), the cost relating to the operation of rail transport services and rolling stock once the network has been deployed can be efficiently provided by more than one company, which can be viewed as actual or potential competitors.

Therefore, from the regulatory point of view, it has been concluded that infrastructure and services can be dealt with in different ways: the former, as a natural monopoly,[5] but also as a potential provider of adequate access to any willing-to-serve operator; the latter, however, can be treated, in principle, 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 existence of asset indivisibilities.

Even though this potential vertical separation can alleviate some of the natural monopoly problems, the rail industry remains a very capital-intensive sector with several other indivisibilities within its productive process. Specifically, the capital units (rolling stock, track and stations) can be expanded only in discrete or indivisible increments, whereas demand may fluctuate in much smaller units. Consequently, increases (decreases) in supply could clearly exceed the increases (decreases) on the demand side, thus resulting in excess capacity. These problems appear in connection to both rolling stock and rail infrastructure. Given that the unit of supply is usually a train or wagon of a given capacity, increases in capacity can only be achieved by the supply of additional units.

This lumpiness of rail transport facilities has several important implications for investment and pricing. For example, the transportation costs of an additional unit of traffic (freight or passengers) may be insignificant when there is idle capacity, but may be substantial when the capital is at the limit of its full use.

Firms can also be forced to employ fixed assets with differing economic lives, whose reliability spans over a large time horizon and affects heterogeneously 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.

Similarly, a final implication of the presence of indivisibilities in the capital assets used in the rail industry 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 even though in most countries it might correspond to an inherited burden from past decades when the structure of traffic was quite likely different from what it is today. Instead, they opt for partial renovations that often introduce technical asymmetries between tracks within a country or region, accentuating their indivisibilities and inflexibilities (Boyer, 1997).

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

1.5. – The organization of railways transport as a public service.

Although not derived from any of its technical characteristics, but rather from historical and organizational reasons, the conception of rail transportation as a public or social service, irrespective of its profitability, is another of the defining elements that have determined the worldwide industry organization and performance during this century. 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 or without accompanying subsidies, public service obligations to rail transport providers in the form of compulsory (often unprofitable) routes, organized timetables or particular services to strategic products or areas. The ultimate reason behind this control, which remains the same today, is that this industry is regarded as an integrative mechanism able 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. – The existence of externalities in the overall transport system.

The policy goal of public service obligation is often supported with the idea that rail transportation contributes less to the rise of negative externalities than other modes of transport, especially road transport. There is abundant empirical evidence showing that the external costs derived from congestion, accidents or environmental impact (noise, visual impact, pollution, etc.) could be reduced if a substantial part of the road traffic market were transferred to the railway sector (Button, 1993).[6]

This intermodal externality arises from the fact that road transport does not fully internalize all the social costs that it generates and economists often recommend the use of congestion and/or pollution rates, for example, to take this into account. However, when these mechanisms are not feasible or politically viable, it might be preferable to lower railway fares in order to obtain an overall improved intermodal balance. These principles should also be considered when defining the appropriate regulation for the rail industry.

In summary, all the above mentioned characteristics – sketched and summarized in Box 1.1 – suggest that an analysis of the regulation of railway transport should be carried out within a general context, taking into account the technological and organizational features of the industry. This study will start with a detailed evaluation of the performance of the industry in recent years.

[6] However, for this to be true in all cases, high levels of demand are also required.

Box 1.1. A summary of the economic characteristics of the rail industry.

Characteristics	Economic consequences for regulation
• **Multiproduct activity**	• Accounting problems. • Coordination of decisions. • Integrated or differentiated management... - between infrastructure and the services? - between different rail services?
• **Structure of rail costs**	• Problems in the definition of rail costs. • Problems in the cost allocation. • Implications on pricing policies.
• **Role of infrastructure**	• Optimum size of railways? • Separation between infrastructure (with characteristics of natural monopoly) and operations (competitive market)? • Access fee to the infrastructure?.
• **Indivisibilities**	• Problems implementing optimal price and service levels. • Dynamic price policies are required. • Investment policies.
• **Public service obligations**	• Financial problems. • Definition of price and service levels.
• **Externalities**	• Implications on (social) optimum prices. • Externality control: accidents, pollutants, energy waste,... • Intermodal implications.

2. – POLICY AND REGULATION IN THE RAIL INDUSTRY.

The overall evolution of rail transportation in recent years as compared to other transport modes is summarized in Box 2.1 for OECD countries. There was a substantial fall of the market share in both freight and passenger markets during the 1970s and 1980s, which apparently stabilized during the 1990s. The decline is particularly relevant because it took place in a period when the total volume carried in both markets experienced a growth of about 50%. Thus, the rail industry appears not to have been able to take advantage of the growing demand for transport in the last 25 years.

This substantial reduction in market share, which is not only particular to OECD countries but also a common trend around the world, can be attributed to both exogenous and endogenous causes. The first category includes the rapid development of alternative modes of transport, especially by road. In the case of passengers, economic growth has fostered the development of the automobile market, which has led to an enormous growth in the indices of motorization. In freight transport, an expanding and competitive trucking sector gained a growing percentage of road transport in many countries. For example, in 1970 in Europe, there were 150 cars per 1000 inhabitants; the figure is now 424. Similarly, the number of heavy vehicles and trucks increased from 7 to 17 million in the 1970-1994 period.

The endogenous causes of the decline can be summarized in the inability of the sector to adapt itself to the changing conditions of its economic environment. Regulation remained obsolete and the rail industry was slow to react. The policies adopted during the

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.

Box 2.1. Market shares of different transport modes (1970-1994).

	Passenger transport				
	1970	**1980**	**1985**	**1991**	**1994**
Rail (%)	10.43	8.64	7.33	6.92	6.85
Private car (%)	77.30	79.97	83.37	84.37	84.38
Bus (%)	12.26	11.38	9.29	8.70	8.75
	Freight transport				
	1970	**1980**	**1985**	**1991**	**1994**
Rail (%)	31.3	23.2	21.2	17.9	15.5
Road (%)	55.2	65.9	69.3	74.0	76.2
Waterways (%)	13.5	10.9	9.5	8.1	7.9

Source: CEMT. *Evolution des Transports. OECD Countries.* (1970-1994).

2.1. – Consequences of the traditional policies on the rail industry.

During the past fifty years, the most common structure of the rail sector in many countries was the existence of a single State-owned firm, entrusted with the unified management of both the infrastructure and the rail 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 regulation.

Due to this protective environment, most national rail companies incurred growing trading 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 that their accounts came to lose all resemblance to reality. Thus, the main problems associated with the traditional policies on railways were: *(i)* increasing losses on the companies' trading account, which were usually financed via public subsidies; *(ii)* a high degree of inefficiency in management; and *(iii)* a business activity oriented exclusively toward production targets, rather than commercial and market targets.[7] These distortions did not come from an artificial reduction in the range of services provided, nor from excessively high fares, but, more commonly, from an unjustified increase in the supply of

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

services (and hence, of costs). Such behavior implies larger public subsidies. The lack, in many cases, of commercially oriented tariffs and investment policies explained many of the difficulties faced by the companies. All these considerations, together with the burden imposed by the technical characteristics of the sector described in Section 1, placed most railways in a very weak position in competition against alternative transport modes. However, this fierce intermodal competition was not, by itself, able to improve the competitiveness of the railway system. As a result, it was necessary to adopt measures affecting the internal behavior and structure of the sector itself. Therefore, the overall decline of the sector sparked a widespread restructuring movement around the world.

2.2. – Recent rail deregulation and privatization experiences.

The worldwide restructuring process of the rail industry began with timid reforms. For example, many countries began by replacing their national railways with autonomous commercial bodies possessing independent and realistic balance-sheets, in which only public service obligations could be explicitly subsidized by the government. Other countries opted for substituting their old geographically based management with a multi-divisional structure, defined by the companies' different lines of business or services. Box 2.2 summarizes these characteristics for several countries allowing us to compare their similarities and differences.

At first glance, one of the common features is that some countries have carried out a relatively long process of restructuring, whereas others have preferred quick implementations over a shorter period. For example, the privatization of the national railways in New Zealand and Japan was phased over several years, while Argentina and United Kingdom did the main parts in less than two years. Another common characteristic is that all restructuring processes were undertaken to make the companies attractive to private investors. Total privatization has been preferred to tendering or concessioning in many cases. This has involved 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. 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, 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 excess staff numbers are two problems that wherever they are found must be dealt with before starting any privatization policy.

Box 2.2.a. Deregulation and privatization experiences in Railways.

	United States		United Kingdom		Argentina		Sweden	
	Before restructuring	**After restructuring**	**Before restructuring**	**After restructuring**	**Before restructuring**	**After restructuring**	**Before restructuring**	**After restructuring**
Market Structure	Competitive situation	Competitive situation with concentration of big companies and many small ones	Public monopoly	Competition for the market. System of 25 franchises in passengers and 2 companies for freight	Public monopoly	Franchise system for 6 freight and 6 passenger operators.	Public monopoly	Monopoly on infra-structure and quasi-monopoly in services
Ownership of railways	Private companies	Private companies	*British Rail* (BR), public body with managerial autonomy	Private concessions and rolling stock leased to private firms,	*Ferrocarriles Argentinos* (FA), public enterprise with little autonomy	Private companies. operating in each franchise	*Statens Järnvägar* (SJ), government department	SJ, public company with wide autonomy, and presence of small companies
Ownership of infrastructure	Owned by railways	Owned by railways	State owned	Private company (*Railtrack*)	State owned	State network open to third parties	State owned	Managed by a public agency, *Banverket* (BV).
Separation between infrastructure and services	Trackage rights exist	Trackage rights exist (Amtrak)	Unified management	Total separation	Unified management under FA	Management of companies. Trackage rights exist	Unified management	Separation. Services run by SJ and small companies. Infrastructure by BV
Regulatory framework	Price control and no closures of loss-making lines	Price freedom and closures of loss-making lines	Freedom of prices, except in some services	Free prices. RPI-X in access pricing.	Prices are regulated	Free prices with maximum level. Minimal frequencies and quality service	Controlled prices	Control over tariffs has been reduced. Not on access prices
Reasons for deregulation	Loss making companies Loss of markets		High level of public subsidy To improve traffic and productivity levels.		High public subsidies, reduce FA's deficits. To improve traffic levels. Improve productivity.		High public subsidies, reduce SJ's deficits. To improve traffic levels. Improve productivity.	

Box 2.2.b Deregulation and privatization experiences in Railways.

	Brazil		Chile		New Zealand		Japan	
	Before Restructuring	**After restructuring**	**Before restructuring**	**After restructuring**	**Before restructuring**	**After restructuring**	**Before restructuring**	**After restructuring**
Market structure	Freight: RFFSA Passenger: CBTU	RFFSA and CBTU divided into sub-networks	EFE (83% of network) and private mining companies (FEPASA)	EFE and subsidiaries, and mining companies	Monopoly in hands of *New Zealand Rail Ltd.* (NZRL)	Monopoly	Monopoly (JNR)	6 passenger companies, 1 freight
Ownership of railways	Public companies	RFFSA is being privatized	EFE, public company. Private mining companies (FEPASA)	EFE and subsidiaries, public corporation and private companies	Public agency	Private (private groups that bid highest to buy the company)	Public	Only 3 in process of privatization
Ownership of infrastructure	Public ownership	Public ownership	EFE and mining Companies	Spread among EFE and other companies.	State owned	Private ownership	State owned	Owned by the 6 new passenger companies
Separation between infrastructure and services	Unified management	Management by the companies. Trackage rights exist.	Unified management of the existing companies	Separation of services and infrastructure on public lines. Trackage rights exist.	Unified management	Unified management	Unified management	Unified management (passenger companies). Trackage rights (4 freight companies)
Regulatory Framework	Regulated prices	Prices control	Regulated prices	Liberalized prices	Prices and service level regulated	Free prices	Regulated prices	Free prices
Reasons for deregulation	Antiquated inefficient railway industry. Reduce State contributions Favor development and regional Equilibrium		Reduction of State subsidies. To improve efficiency of system Increase market share		High public subsidies and reduce NZRL's deficits		Reduction of State subsidies. Improvement of the productivity	

2.3. – Vertical separation and private participation in the rail industry.

One of the most clearly defined patterns that emerges from deregulation and restructuring measures is found in the extent to which they carry out two critical dimensions summarized in Box 2.3: the degree of separation between infrastructure and services, and the involvement of private management in the sector (Kopicki and Thompson, 1995).

2.3.1. – Vertical separation.

With respect to the first dimension, there are three main different options for the vertical organization of the railway industry: *(i)* vertical integration, *(ii)* competitive access, and *(iii)* vertical separation. The first structure corresponds to the traditional and historical 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.

Competitive access is characterized by the existence of an integrated operator (usually public), which is required to make its rail facilities (tracks, stations, etc...) available to other operators on a fair and equal basis through the trading of, for example, circulation rights. This method keeps the advantages of integration (in terms of economies of scope, coordinated planning and reduction in transaction costs) but, as will be discussed in Section 5, if the integrated company has incentives to leave out other operators, the overall effectiveness of the system may be doubtful.

Alternatively, in the complete vertical separation scenario, the ownership of facilities is fully separated from other rail functions (train operations, marketing, etc.) and can also be privatized. This form of restructuring is very attractive because rail infrastructure, which remains characterized by natural monopoly conditions, is separated from rail operations, where potential competition among different operators may be implemented. In general, the main advantage of this vertical unbundling is that rail transport is placed in a similar situation as road transport, especially in regard to infrastructure planning and the tariff system. Hence, governments could study investment proposals on the basis of a cost-benefit analysis, while pricing policies could be based on the criterion of social cost. An important problem here lies in the difficulty of defining the social cost of railway infrastructure use. The determination of 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, the quantification of this opportunity cost is difficult, especially if there is a mixture of social and commercial services.

In addition, separation of infrastructure from services greatly facilitates the entry of more than one operator on a single route. In profitable services or routes this system would permit notable improvements in the efficiency of the industry by allowing direct competition among operators, and thus eliminating standard monopolistic practices in the sector. In non-profitable services, infrastructure separation can be accompanied by

tendering, thus stimulating increased efficiency through competition for the market, the introduction of innovations, and a clear improvement in marketing.

However, the vertical unbundling of the rail industry also implies several disadvantages that must be evaluated. The main problem is the potential loss of economies of scope derived from the joint operation of tracks and services. It is often pointed out that the relationship between the services supplied and the rolling stock used, as well as the quality, quantity and technical characteristics of the infrastructure, is so close that both aspects need to be planned together. Thus, the assignment of different services to several operators may imply a lower utilization of the staff and physical assets of the sector. Another negative argument is based on the risk that the new system may be less attractive to the user than an integrated system.[8] The separated system also facilitates the growth of transaction costs and of a complex institutional framework. This is the argument that, in order to bring about a degree of competition in rail transport, the government must postulate such a complex institutional arrangement that each organization will be involved in negotiating a huge number of contracts, with the result that transaction costs may become prohibitive.

Finally, one last disadvantage of the process of vertical separation of infrastructure and services is the reduction of investment incentives by the company or body managing it. 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 the higher the competition in the market, since competition weakens the bargaining position of individual operators in negotiating with the infrastructure owner.

2.3.2. – 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 rail industry. The first situation is a *government department*, where the railroad is fully controlled by the government or the relevant Transport Ministry, so its degree of independence is zero. Ownership and operations are fully public and financed by budgetary transfers.

The second example is a *public enterprise*, where the railway is characterized by a higher managerial autonomy, although still subject to government approval for many of their decisions. Normally, these railways sign more or less formal *contracts* with the government, where the objectives and responsibilities of each entity are clearly specified. In addition, it is usual that some restrictions on the levels of public subsidies are gradually introduced. Similarly, the case of a *reformed public enterprise* corresponds to a situation where the railway is corporatized (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, decides the pricing policies and the

[8] For example, because of the failure to achieve even headways, or the lack of interchangeable ticketing or the lack of an integrated information system).

investment levels, while guaranteeing, by the necessary subsidies, the supply of non-economical social services.

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

Box 2.3. Alternative organizational structures in railways.

		Vertical Unbundling		
		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 (in progress), US (Class I), Canada	UK (freight, infrastructure)

Source: Elaborated from Galenson and Thompson (1993).

Concessions are a broader form of lease in which the contractor also agrees to make certain fixed investments and maintains the use of the assets for a longer contract period. At present, it is the most used way of restructuring the rail industry, and it will be extensively discussed in the remaining sections of this paper.[9] Finally, there are *joint ventures*, which entail the largest degree of private participation, where private partners contribute development capital, planning and management expertise to develop land or other real estate owned by a railway. There is also full *private ownership*, where certain services or whole companies are operated by private firms. Box 2.3 summarizes the overall structure of the railway industry around the world in these two dimensions.

2.4. – New regulatory scenarios in the rail industry.

Present circumstances, as well as the importance of the vertical separation/private participation dimensions, creates a new framework for increasing the liberalization of the rail sector. This significantly introduces new roles and functions for the regulator and modifies the number of possible regulatory structures. The choice of each particular method for railway restructuring will actually depend on a number of particular objectives or goals that the regulator must clearly balance and weight accordingly within the economic environment in which its activity takes place.

Box 2.4. Different rail regulatory scenarios and their objectives.

Scenario	Objectives						
	Fiscal	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.	✓	✓	×		×	×	×

In particular, one of the first elements to consider is the existence of *financial constraints*, in terms of the maximization of the proceedings obtained by the State after the restructuring process and the subsequent minimization of the State's financial burden. 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

[9] In fact, the exact denomination of the privatization agreement, either *concession*, *license* or *franchise*, for example, varies among countries and depends on the final property of the assets.

resources. Similarly, there is the goal of attaining *allocative efficiency* by setting optimal prices equal to the marginal social cost, which, from an inter-modal viewpoint, facilitates the best distribution of traffics. The objective of *dynamic efficiency* requires the minimization of costs in the long term by means of active and technology-improving investment policies. There can also be *equity objectives*, such as facilitating transport for any citizen independent of his level of income. Finally, the government can also consider the goal of optimal *allocation of capacity*, which favors the management of railway capacity as well as the co-ordination with other modes of transport and the overall *minimization of risks* in terms of maintenance of the service across time, risk of default, etc. The combination of these objectives creates at least eight different possible regulatory scenarios grouped in decreasing order of private participation (see Box 2.4.). Some other possible scenarios have been eliminated (such as those relating to the mixed forms described above).[10]

The objectives enumerated in this box could be given a different weight. For example, financial and cost efficiency objectives have recently been valued above all other types, which explains the boom in privatization measures, both by a system of concessions and by direct sale to the private sector. In addition, there is an evident tradeoff between the financial and efficiency objectives and those of a social nature as the degree of privatization of the scenario increases. Thus, the scenarios characterized by the existence of a public company clearly serve social objectives (equity, reduction of risks on the service, intermodal co-ordination, etc). This type of company, however, was shown to be inefficient and likely to produce huge commercial deficits. As we have already indicated, these were the main reasons for the restructuring of the sector.

The deregulation measures that define Scenarios 4 and 5 (concession system) have the advantages of favoring the efficiency and solvency of the companies, as well as reducing the State's financial burden (although possibly these effects are not as great as in 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 quality levels, so that impact on equity can be minimized. Likewise many routes can continue to be served which, though not profitable, are beneficial from a social viewpoint. The concession of these routes to the operators who request a lower level of public subsidy meets both efficiency and equity objectives.

In regard to dynamic efficiency, the first results of the levels of investment implemented by the restructured companies or bodies are ambiguous. Thus, 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 existing in the

[10] This is because many of these forms of private participation are related to very specific services (in 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.

previous regulated context. In this sense, other experiences such as those of Brazil have indeed led to a substantial recovery in investment levels in both infrastructure and rolling stock, with an evident 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).

Operational risks, apart from other types of considerations, are minimized when a public enterprise is entrusted with their management. With a private company, there is obviously 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, there is a problem associated with the management of capacity. This problem is easily eliminated in the case of vertically integrated companies.

On the other hand, the problem is not so simple for systems of competitive access or of separation. In this case, the problem is increased for companies with high traffic densities and conflicting demands for capacity. It is true that modern computer technology can reduce the problem through the real time management of electronic systems. However, when connecting systems have different dispatching priorities and differing informational qualities, it will be very difficult for any person or railway to plan and manage integrated services across several systems.

2.5. – Design of concession contracts for railways.

In spite of the number of potential regulatory scenarios just described, few railways around the world have been truly privatized. Instead, most countries have opted for concessioning their rail services and, in some cases, even their rail infrastructures, to private firms in exchange for a fixed payment.[11] This restructuring form has been favored because it allows the government to retain ultimate control over the assets while the private sector carries out day-to-day operations according to some pre-specified rules devised in a contract (Thompson and Budin, 1997), from which the problems associated to traditional regulation are transformed into issues of contract enforcement. Since the number of variables to consider is large, rail concession contracts cannot be reduced, in general, into a single standard model. However, according to the existing experiences, Box 2.5 proposes six key variables to consider.

The first critical aspect to determine in a concession is its type, both in its vertical (or functional) size and its horizontal (or geographical) one, according to the features already described in Section 2.3. The most often used approach in recent concessions in the rail industry has been to create smaller packages attending to the horizontal dimension of the industry in the country. For example, rail freight systems in Argentina, Brazil, Mexico or 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 these countries have also used economic criteria in designing the

[11] 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.

size of the concessioning package, taking into account the profitability of different lines. In Europe, functional separation between infrastructure and services has been preferred instead, particularly since the European Commission Directive 91/440 and its follow-on orders. In its most extreme degree, this form of concessioning of rail services was used in the privatization of *British Rail*, which also included the private provision and management of rail infrastructures. A less extensive vertical separation approach has been developed in Sweden and other European countries, where the infrastructure has not been auctioned off to private firms (Lundberg, 1996).

Box 2.5. Key variables in designing rail concession contracts.

Type of contract	• Package size depends on economies of scale/scope and existing potential for competition • Horizontal concessions (geographic) according to country's characteristics • Vertical concessions (functional) according to network's characteristics (including current state of infrastructure and new investment needed) • Mixed packages depending on profitability and bidders' financial constraints • Freight vs. passenger concessions depending on relative traffic shares
Award and duration	• Pre-qualification requirements to reduce risks • Type of auction (sealed, one-shot) and explicit rules for auctioning • Selection based on government's objectives (fiscal, equity or efficiency) • Short periods (favor competition; diminish investment incentives) versus long periods (favor investment; diminish enforceability) • Termination: re-auction preferable to automatic renewal

Contents	• Concessionaire:
	  • obligations: services (with adequate performance) and payments
	  • rights: exclusivity and compensation for public service obligations
	• Regulator:
	  • risk sharing (net cost/gross cost mechanisms)
	  • asset ownership

Price control	• Price control depending on monopoly power and social objectives • Ideal criterion: marginal cost rules • Practical mechanisms: rate of return regulation and price cap schemes • Other schemes: price discrimination and cross-subsidization

Quality regulation	• Definition of quality targets in the rail industry
	  • Quality of service
	  • Safety and externalities
	  • Dynamic quality: investments
	• Instruments for quality control

Infrastructures	• Access to rail infrastructures • Access pricing • Coordination and intermodal competition

The second key issue in designing the contract for concessioning rail services and infrastructures is correctly defining the award process and the duration of the concession. If precision in procurement and sustainability against legal challenge is to be achieved in the design of concession contracts for the rail industry, common sense suggests that everything in the contract should be unambiguously defined. This includes a number of dimensions, such as the rules for the auction and the criteria defining how each concession will be awarded to a private company or a consortium of companies. In practice, there are a

number of elements to choose among, such as, for example, the award criterion between maximum payment to government (or minimum payment by government, in negative concessions) and minimum tariff. There can also be a choice between unrestricted bidding and bidding that involves some sort of pre-qualification, together with the exact choice of the type of auction (see Guislain and Kerf, 1995, and Kerf *et al.*, 1997).

In the privatization of former *British Rail*, for example, the process of concessioning began with a pre-qualification stage, followed by a formal invitation to tender for a particular package. After indicative bids had been received, a short-listing process occurred, usually with four bidders remaining. One of these was subsequently named as the preferred bidder, and was given around a fortnight to complete financing and other organizational arrangements before being confirmed as a winner. At that point, the regulator gave public details of the bid, in terms of the subsidy required and the service improvements promised.

With respect to the bidding mechanisms, there is an extensive literature and experience on different auction forms. Single, sealed enveloped bids are the simplest form to avoid collusion and obtain higher bids, but more complex approaches, such as real time auctions, have been used in some transport concessions (for example, the Brazilian freight railways). Once the rules have been setup and bids submission requested, bidders should have a period of study in order to form their own valuations according to the potential gains to be extracted from the concession. Early research by Preston *et al.* (1996) for United Kingdom indicated that their key issues for the bidders were the length of franchises, the level of competition that would face from other operators, the separation of infrastructure from services and the costs (including new investments) associated to their maintenance and the own mechanisms and selecting criteria of 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 not a single method to select the winner once the bids have been submitted, and the final choice will depend on the regulator's objectives, which should be explicit and based on transparent criteria. Thus, if the government intends that private participation should be a means of reducing the burden on the public purse for the provision of rail services, it must use fiscal benefits as the greatest criterion, and it will look at who needs the lowest subsidy or who offers the highest auction price. In Brazil, for example, the six regional rail concessions were successfully auctioned to highest bid above the government's stipulated minimum price. Concessionaires were required to make an up-front payment immediately after the auction and then 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 for winning bidders, especially in the first concessions. Other criteria were the financial position of the tenderer, its managerial competence and its operational proposals. In any case, regular payments or subsidies can be made dependent on the firm's performance, so as to maintain a degree of flexibility that favor the development of the industry (Preston, 1996).

Alternatively, if the regulator sets the tariffs and defines the quality of service in the contract, bids can be evaluated on the basis of the lower cost provider whilst simultaneously including penalties for the case of not achieving certain performance objectives. These social objectives can be also targeted by focusing on the bids which

propose to monopolize the industry the lowest number of years or to charge the lowest fare to final users, provided this fare can be properly defined. 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 later possibilities of regulatory interventions on prices and demands an adequate definition of quality standards.

Many concessions in the rail industry have been awarded using a multicriteria points formulae, where it is possible to take into account a larger number of objectives. This method was used in Argentina, where 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 weighting of these criteria reflected the importance attributed to investment in the railways, but also political compromises on employment.

The rule of keeping things simpler was later adopted for the award of metropolitan commuter railways in order to make the bidding process and the final selection as transparent as possible and not to induce exaggerated or misleading assumptions. Thus, the Argentinean authorities learned from the freight concession experience that trying to select the winning bid through numerous cumbersome criteria with discretional 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 and known to all potential bidders and bidding should take place on the basis of a single parameter, which encompassed the economic assumptions made by the bidders in relation to the terms of the concession. In the case of the metropolitan railways 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).

With regard to the optimal duration of the concession contract, the tradeoff is evident in terms of efficiency, since the shorter the concession, the more immediate the competitive pressure, but the less 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 objectives of the regulator. This balance often creates conflict. Thus, while concessionaires argue in general for long contracts on the grounds that this provides them with incentives to build up the business and purchase or replace long-lived assets, regulators prefer shorter lengths to favor the achievement of efficiency (by the implicit threat of not renewal) and fiscal goals (since the canon or auction price may be increased after the first years of the concession). Only if sunk investments are minimal and asset re-utilization 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 rail concessions is about five or ten years when they refer only to services and up to thirty when

network investment and development are included. The figures vary from country to country. In Argentina, for example, the six freight packages created were concessioned on a thirty-year term, with an optional ten-year extension, due to the poor state of infrastructures and the huge investment that was required. The international rail link between Côte d'Ivoire and Burkina Faso was awarded in a fifteen-year concession for similar reasons. Conversely, the train-operating companies in the United Kingdom were granted a concession to run passenger services for a period of only between seven to fifteen years.

After the duration period has expired, the contract must also specify several termination arrangements in order to avoid any disruption to railway services. A first 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 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. An example of this strategy was given by 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 end in deteriorated performance, the concession contract should specify the circumstances under which renegotiation is possible and desirable, and which party should initiate the process. A pre-scheduled revision process, if certain intermediate objectives are achieved, might help to reduce both parties' risks. Although the contract will always be incomplete, standard clauses should include the behavior under unforeseen changes in demand conditions, the responses to not anticipated rises in energy or labor costs, etc. In Argentina, freight concessions, for example, could not fulfill their promise to invest about $1.2 billion in the rail network over fifteen years due to unexpected falling traffic levels.

In all cases, a flexible mechanism for contract renegotiation is required, since the regulator sometimes faces the dilemma of enforcing contracts to the detriment of the operating companies and the national rail system, or he might have to reschedule investment and make other compromises at the cost of undermining his credibility for enforcing the agreements in the future (Carbajo and Estache, 1996).

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 winning the concession 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 certain exclusivity rights that (totally or partially) protect her from other rail competitors.

On the other hand, in exchange for the payment or the compensating subsidy, the regulator promises to monitor the overall performance of the sector and provide a stable framework for current and future rail operations, which may include infrastructure provisions if this was auctioned off to private firms. In fact, a large part of railway activities might be concessioned. These include, in infrastructure: track, signals and stations, yards, shops; in operating equipment: locomotives, wagons, carriages; and in general services access to track, route and schedule information, and maintenance. The exact form in which

this process is developed in practice mostly depends on the risk sharing agreements achieved by the parties.

According to a service contract, for example, the train operators provide rail transport services for passengers or (rarely) freight that are fully specified by the regulatory authority, in terms of routes to be served and the minimum quality and technical requirements. The operators may also cover some investment costs and carry some commercial risk. This risk can be integrated into a net cost contract, where the operator keeps all the revenues generated by the 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; thus, operators carry cost risk, but no 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 fares scheme, with modal integration and quality control.

The regulator may retain control over and responsibility for common functions, and its main role should be mostly restricted to regulating quality (in terms of service, safety, environmental and technical standards), to control monopolistic behavior (in terms of abusive prices or services), and to determine the overall characteristics of the functioning of the sector (in terms of coordination at the national and international level) under the competition rules or rights established for all firms as well as antitrust and commercial legislation.

The form in which the exclusivity rights associated to rail concessions have been implemented varies in each country. In Argentina, the rail freight concessionaires have exclusive use of the 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, whereas freight concessions were awarded to private firms competing with each other. Particularly, in the Côte d'Ivoire-Burkina Faso trans-national railway, the fifteen-year concession was awarded with a single seven-year exclusivity period, after which the operator must grant track access, for an agreed fee, to any third-party carrier specified by the regulator.

Thus, exclusivity rights should be viewed as another instrument of regulatory control, and not taken as granted by the firms *ex-ante*. By limiting the duration of the monopoly period, a balance is struck between the regulator's desire to reap the benefits of allowing competitive access to the tracks, and the private train operators' preferences for full control over the market in order to generate profits and make forecasting revenues easier. In general, most railways have been concessioned on an exclusive basis in geographical areas, as in Argentina or Brazil, with possibly 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 for promoting competition in major markets (as in Mexico) or for non-competing services

(such as passenger services on freight tracks in Chile) according to the traffic mix in each case.

With respect to the concessionaires' obligations, the private provision of rail transport services, particularly when they pertain to less developed areas or zones with historical structural deficits in track network, cannot always be separated from public subsidization or compensation in reciprocity for public service obligations imposed with political criteria. In these cases, the concession contract for loss-making but socially necessary services must clearly include these arrangements, in terms of detailed performance levels to be attained by the firm, and may even be designed in order to be awarded to the company willing to provide the specified services for the lowest level of subsidy (negative concessions), as in Argentina.

As a final feature in defining the rights and obligations of the concessionaires, the current experience of rail concessions in South America show that the restructuring process has often lowered employment levels. Although, in principle, this should not be a relevant feature for operating the sector once it is privatized, in practice it becomes one of the most difficult obstacles to hinder the private participation of rail operators in certain countries. In Brazil, for example, big redundancies were inevitable and were taken 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 redundant staff not hired by the concessionaire. From then on, compensation for any additional employees laid off is the responsibility of the private operator. Undoubtedly, any constraint to the service, either on employment or on other productive factors will be reflected in the auction price of the concession.

In summary, in their general form, rail concessions define the most advantageous form to solve the challenges posed by the new regulatory environment in which the rail industry is currently placed. It usually adopts the form of a long and medium term contract in which 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 remain three other particular aspects of this concession contract design that, by their own importance in the regulation of the rail industry, deserve particular treatment in the following sections. First, in Section 3, price regulation, in terms of defining the most important issues to establish effective and well-oriented price control mechanisms within a concession contract. Second, quality regulation, studied in Section 4 and including both its static dimension (quality of service, safety and environmental issues) and dynamic dimension (rules for infrastructure investment and financing). Finally, Section 5 will be devoted to those aspects related to the coordination between infrastructure and superstructure in the rail industry.

3. – PRICE REGULATION IN THE RAIL INDUSTRY.

According to well-established economic principles, in order to make the most efficient use of the resources available in an economy, the price charged for any rail transport service should match the opportunity cost of its provision. This is the economic efficiency or *first*

best criterion around which traditional regulation of the rail industry has been defined in the past. Thus, the main focus of government regulation in this industry was the control of market power through regulated prices envisaged to limit the degree of monopolistic abuse of a particular railroad. The exact form of tariff control (official approval of rates with little or null degree of financial autonomy) depended in each case on the nature of the sector, the ownership of the assets, the complexity of the regulated service, and the social and political pressures to keep the firm's financial equilibrium in the medium and long run.

In practice, however, this opportunity cost pricing rule entailed huge measurement difficulties and, in particular, in industries with large economies of scale such as the railways, yielded uneconomic losses (Amstrong, Cowan and Vickers, 1994), as described in Section 1. Therefore, this form of regulation was complemented by a number of standard price mechanisms devised by economic theory to substitute the ideal efficiency criterion of pricing each unit of service at the exact cost of its provision.

Price discrimination policies, for example, were, and still are, common in transport, either by type (students or elderly tariffs, frequent or commuter travelcards), number of consumers (group discounts), type or volume of freight (cargo rebates for some goods) or by time in the day or season (peak-load prices). The use of two-part tariffs, with one fixed component and a variable one, is also a common tariff policy in which each unit of consumption (for example, a single trip) is priced differently. All these mechanisms allow for a greater flexibility for the railways and increase their revenues without much affecting their costs, but both their social acceptability and the informational requirements they demand can limit the extent of its application.

In the new regulatory environment defined in the previous section, in which separation of the infrastructure from services can be relatively easily achieved, and in which a notable degree of private participation in rail management exists through concession contracts, pricing principles must be put in practice by means of concrete rules within the contract.

Since rail concessionaires are now able to set prices more or less freely, the concession contract should include a procedure to control the prices set by operators. These price control mechanisms should be set, in general, according to three key factors: *(i)* the degree of monopoly power effectively conferred to the operator; *(ii)* the extent of government non-commercial objectives previously defined in the concession award procedure; and (*iii*) the possible existence of limiting factors, such as intermodal competition. This latter element is particularly relevant in the case of rail freight operations (intermodal competition from trucking),[12] but in the case of passenger traffic (particularly the commuter and regional one) social pressures to keep fares low usually dominate many price interventions. In practice, the most common alternatives (*second best criteria*) to control prices in rail concessions are rate of return regulation and price cap mechanisms.

[12] For example, in Argentina railways only carried 8% of total freight tons-km at the time of concessioning. In Chile, where the average haul distance does not favor rail.

3.1. – Rate of return mechanisms.

Rate of return regulation is particularly used in railroads in Canada, Japan and the United States. The principle behind this type of regulation is to constrain prices so that the regulated rail transport operator earns only a fair rate of return on its capital investment. Under rate of return regulation, the regulator typically determines a revenue requirement based on a firm's accounting total costs during a test year, according to the variable costs and an estimate of the cost of capital to the firm, given by a "reasonable" rate level multiplied by a rate base (Liston, 1997).

$$\text{Revenue requirement} = \text{Total Cost} = (\text{Variable Cost}) + (\text{Rate Level} \times \text{Rate Base})$$

Thus, rate of return regulation has three components: the rate base, the allowed rate level, and the rate structure. The rate base refers to the investments that are allowed to earn a rate of return; the rate level refers to the relation of overall revenues to costs, and the rate structure determines how individual prices are set for different services or customers. Determining the first of these three components is often the most important regulatory task under this form of regulation, since inadequate calculations of the rate base might either jeopardize the survival of the firm or allow it to earn excessive profits. In practice, the rate base usually includes most fixed costs less depreciation and working capital.

Three characteristics should govern the definition of the asset rate base affected by the regulatory activity. First, with respect to the treatment of past investments carried out by the railroad before the regulatory period,[13] it should be consistent and transparent in order to ensure that assets are not expropriated *ex post* by an opportunistic regulatory behavior, which would increase the cost of capital required by investors. Second, with regard to future investments, and their associated expected operating expenditures and costs, they should be considered into the asset base definition inasmuch as they do not imply "excessive" investment and only when they are fully incorporated to the firm. Finally, with respect to current investments, the problem lies in determining the value of the firm's capital. If the existing assets were transferable to other activities without cost, then the conceptual problem of determining their value would be simple: their replacement cost or resale value. At the other extreme, and more frequent in the rail industry, is the case where existing assets are sunk, so that the opportunity cost of using them in their present activity is zero. Then, if the regulator seeks maximal efficiency, it should ensure that rate of return structure (and, indirectly, prices) are set to cover future avoidable costs.

Since most of the assets currently used by railways are sunk and financed before the concessioning process, both of these solutions are troublesome. Market values are much lower than replacement costs so that this valuation would yield large price increases and windfall gains for private stakeholders at the expense of consumers. On the other hand, attributing a zero value to the existing assets would make the windfall gains go in the

[13] This is often the case in many restructuring processes, when a former state-participated railway has transferred its assets to private concessionaires.

opposite direction and the actual proprietors would be reluctant to finance future investments with such a lower real return. A tentative method to address this problem is to use some average procedure that consider either a financial projection of what will happen with the future rate base or calculate indicative values by estimating the cash-flows that the firm would have earned had the regulatory regime remain unchanged.

Despite its numerous advantages within the traditional price regulation mechanisms (mainly in terms of simplicity), three additional problems are associated with this sort of regulation. First, there is little incentive for productive efficiency, since firms can pass production costs onto final users in the form of higher prices; second, it leads to excessive investment and capital use because the firm is guaranteed a return on investment;[14] and, finally, the high degree of discretion usually enjoyed by the regulator in determining the rate base and the rate of return reduces incentives for rent-seeking behavior on the part of the regulated firm.

3.2. – Price cap regulation mechanisms

The most common alternative to the standard rate of return regulation is the use of cost-plus incentives that, in practice, take the form of a menu of costs reimbursement rules in which firms would self-select themselves according to their preferences for sharing the operating costs with the regulator.[15] The basic aim of these mechanisms is to favor the achievement of dynamic efficiency (in the sense of the regulated firm achieving the lowest unit cost in the long run) by sharing some of the rents due to efficiency improvements between the firm and the regulator.

There are several ways to accomplish this goal, and implement its results. For example, the sliding scale plans used in the United Kingdom *Railtrack*'s regulation consist of a price adjustment mechanism through which the actual rate of return earned by the firm is adapted to changes in productivity according to a variable parameter.

Price-cap regulation is another incentive method used both in railways and other privatized utilities. In its most standard form, it consists in setting traditional maximum price schemes designed according to long-run marginal costs in order to offer a firm an incentive to achieve the goal of dynamic efficiency while maintaining all or part of the gains associated with the future increases of efficiency by the firm. This mechanism was born as a consequence of the criticism directed at the lack of cost minimization embedded in rate of return regulation and other traditional price regulation mechanisms. However, its efficiency gains have to be balanced with the higher informational rents that it implies (De Rus, 1998).

There are a number of minor variations of the price-cap system. One of the most developed ones in the rail industry is the RPI-X formula. In this setup, the price for a basket of the firm prices can increase in any one year by no more than the increase in the retail

[14] This is the so-called Averch-Johnson or capital-bias effect, which is not particularly adverse in less developed economies whose capital needs are seldom fulfilled.

[15] See Guasch and Spiller (1996) for detailed examples in other industries.

price index (RPI) for that year, minus some fixed-cost (efficiency related) parameter X.

$$\text{Price}_{(\text{year } 1)} \leq \text{Price}_{(\text{year } 0)} \times (\text{RPI} - \text{X})$$

In the case of multiproduct activities, this expression can be easily adapted by requiring that a certain weighted average of percentage price increases not exceed the rate of growth of the RPI less X percent. The weight for each price can be defined according to the share in total revenue of each product or, alternatively, it can be imposed that the average revenue (calculating with accounting figures) can grow by at most RPI-X. Thus, the regulator can control the prices of multiproduct firms by focusing on their revenues, correcting them according to adequate weights. It sets the price for a certain number of years starting with a reference price often calculated according to rate of return criteria.

In the United Kingdom, for example the price cap mechanism, in its RPI-X formulae, has been applied to passengers traffic franchisees.[16] Commuter fares are regulated with respect to a basket containing all relevant fares, weighed broadly by the income that the operator derives from each. From three years from January 1996, increases in the capped fares are not permitted to be more than the retail price index increase from the 1995 base price; for the years from January 1999 the price cap is planned at RPI-1%.

The purpose of this method is to increase the efficiency of the regulated rail operator, allowing the firm to earn substantial profits by improving their efficiency while simultaneously financing current and future operations. This implies that, in practice, when setting the level of a price cap, the rail regulator must consider several factors: the cost of capital, the value of the existing assets, the future investment programs, the expected changes in productivity, estimates of demand growth, and, perhaps, the effect of X on actual and potential competitors. Some of these are common to other price-regulation mechanisms and, in particular, they are needed when using rate of return regulation, as described above.

There are different procedures and rules to deal with each of them. The cost of capital and the value of existing assets can be calculated using standard financial techniques. The future investment program and its implications depends on both expected changes in productivity and estimated demand that can be obtained from econometric techniques or simpler projection and analysis of historical data. Finally, the effect of the price-cap on the future shape of the market must be conjectured from past experiences or yardstick comparisons.

One of the most critical issues is the setting and resetting of the productivity X-factor. A possible method consists in using indexes or indicators (as those described in Section 6) to measure the difference between aggregate rates of growth of outputs and inputs and therefore by the residual calculate productivity. Again, econometrics can provide alternatives for estimating cost functions and their corresponding productivity parameters

[16] RPI-X is also used for access pricing, as will be described inSection 5.

(see Borts, 1960, for a classical reference). Once the X-factor is determined, the choice of initial price ceilings, which are first imposed on the firm after a switch of regime, is critical. If the caps are too high, then too little of the surplus is transferred to consumers and deadweight losses are huge. If they are set too low, the firm may be unable to break even. It may then have difficulty in attracting capital, and its service quality may deteriorate.

Another important element of RPI-X regulation is the existence of cost passthrough provisions, through which the firm can translate to their customers unexpected increases in certain factors outside its control. Although these clauses are standard in the regulation of other utilities, they are not frequent in the rail industry. The most plausible case could be given by energy costs, for which a certain percentage (100% or less) of the cost passthrough onto the customers could be established in the concession contract.

In summary, the pricing principles traditionally used in the rail industry are not particularly different from standard economic principles. On the contrary, they have been extensively used as examples for other transport modes or economic sectors. Both rate of return regulation and price-cap mechanisms are the most commonly used price-regulation schemes in the rail industry nowadays, and they represent a form of price control in which, as opposed to the traditional regulation, some commercial freedom is given to the regulated firm. Although rarely implemented in their purest forms, rate of return regulation and price caps (in its most developed form of RPI-X) center most of the debate and practical experiences in rail concessions.

These methods are valid not only for limiting monopoly profits earned in passengers or freight traffics, but also in the control of infrastructure access prices, as will be shown in Section 5. Finally, since tariffs control could easily be cheated on quality grounds, quality requirements become essential for monitoring the overall performance of the rail concessionaire. This is precisely the objective of the following section.

4. – QUALITY REGULATION IN THE RAIL INDUSTRY.

Quality performance is not neutral for the economic contribution of the rail transport sector to the social welfare. The particular level of quality achievement of the train operators and their particular features in regard to the three main dimensions that broadly define quality in the rail industry (service, externalities and investment) critically determine the value added by this transport mode. The first question that naturally arises is why quality regulation is needed at all in this industry, and to what extent this regulation relates to standard price regulation mechanisms described in previous sections. Economic theory provides a well-known argument to answer these questions: real world transport activities are characterized by market failures due to informational problems.[17]

In an ideal world with a large number of competitive rail transport service providers and well-informed consumers of passenger and freight transport services, quality regulation would not be required since market forces would gradually adjust the consumers' wishes or

[17] There exists an extensive literature on this topic. See, for example, Stigler (1987) and references herein.

demands (in terms of prices, levels of output and of quality of service) to the firms' supplies. Thus, if no price correction takes place, less punctual or unreliable rail companies, or those with poorer freight management performance, would be driven out the market and only those whose price-quality ratios were in accordance with demand would remain. However, when full information no longer exists, markets cannot exert this disciplinary role on the firms and purely competitive solutions are not always positive in terms of quality (or price or output). Unsafe, unreliable or unpleasant services may result under pure competition of rail transport operators, since the limited availability of adequate resources and the lack of adequate control mechanisms make it impossible to adjust consumers' and producers' interests.

In the case of the traditional organization of the rail industry some years ago, as a monopolistic structure with a single firm providing services at the national or local level, the price-quality adjustment problems may even increase, since the monopoly's privately optimal level of quality may not coincide with social standards in many cases. Simple price regulation seldom suffices to solve this question. Any regulated multiproduct monopolist in an environment of asymmetric information tends to degrade qualities in order to achieve a larger level of profits once it has entered the market. Railway firms are not immune to this temptation, with quality of service specified, for example, in terms of punctuality and cancellation standards. In particular, and not just for the rail sector, the quality outcome of any monopolist heavily depends on the specific method of regulation adopted. For example, under a rate of return regulation, overinvesting in non-required technological quality may accentuate the Averch-Johnson effect. Alternatively, under a price-cap regulation, a subtle cut in quality can be a very tempting way of cutting costs (Carbajo, Estache and Kennedy, 1997).

Therefore, the price regulation mechanisms analyzed in Section 4 should be considered incomplete if they do not include quality provisions. This is not always an easy task in practice, since adjusting those price mechanisms by quality may result in making them inoperative or excessively difficult to manage by the firm or to monitor by an external regulator. For this reason, instead of correcting price control mechanisms, most regulators determine the quality standards or targets to be accomplished by train operators.

4.1. – Definition of quality targets in the rail industry.

In setting up those quality standards incorporated in the design of concession contracts, the regulator often makes use of the principles envisaged in yardstick competition.[18] These quality standards may be constructed at the national or regional level with inter-industry comparisons (as in Brazil or Chile with regard to many of their public utilities) or by establishing international benchmarks or best practices (as in Australia with regard to its transport services and infrastructures).

[18] To avoid the problem of regulator's capture and the discretionality of the regulatory action. However, there is a risk of making undue comparisons between different rail systems.

Three elements are considered in detail when designing this process. First, as in other transport modes, quality is mainly measured in concrete service levels or specified service standards. However, this measurement is suited more for factors such as the punctuality of trains, the reliability of aboard services and the waiting time at stations or platforms, than it is for other factors.[19] Simultaneously, the services provided before the transport itself, such as ticketage, reservations, or luggage or cargo handling at stations or wagons are often ignored as part of the value chain of the railway industry, although they may constitute relevant aspects of both intramodal and intermodal competition. For these reasons, the first element to consider in designing a quality control in the rail industry is an integrated vision of the transport service that includes not only the ride itself, but also all aspects related to infrastructure (track and stations), stations and pre-transport and post-transport services provided to clients.

Secondly, a further aspect of quality regulation with particular relevance for the railways, is the flexibility with which scheduled services can be changed, or new services introduced in response to changes in the levels or patterns of demand. The rail industry has always been at a disadvantage to roads and air in this respect because of the need to coordinate working timetables and operations within certain technical requirements characterized by a lack of alternative routes between two points.

Hence, it is not very easy (with a few number of increasing exceptions in many countries) for rail transport to offer on-demand services to passengers (for example, as done by charter airlines) or cargo customers (with door-to-door services) unless the rail network allows it. Coordination is thus particularly relevant for quality of service regulation within the rail firm itself, and must also be considered when designing the industry structure. For example, one potential disadvantage of the split between infrastructure and operations is that this might become even more difficult when such changes have to be negotiated between different organizations, especially where approval of timetable changes also needs to be secured from other train operators using conflicting train paths.

With regard to other industries, intermodal coordination is also needed, since the evaluation of social quality performance is always done in relation to feasible alternatives. Saturated corridors (where the investments in roads, railways and airports clearly overcome the demand) are a waste of resources that few economies can assume. This almost general equilibrium approach to evaluating quality constitutes the third element of the quality regulation process, although, in this case, this is not particular to this industry. The socio-political implications of quality regulation (in terms of equity or public service obligations and the social acceptance of quality standards) determine at last the overall quality targets to be established in each industry.

[19] For example, railway track can deteriorate with respect to the smoothness of the ride provided, or the noise or vibration generated to passengers and third parties (buildings close to the tracks), even though punctuality and or safety are not jeopardized, so there might be incentives to reduce maintenance standards in this respect.

Box 4.1. Quality dimensions of the rail industry.

Dimension		Definition	Measurement Variables

Quality of Service	**Vehicle**	Aboard quality (wagons, locomotives)	- Age of vehicle/number of years in service
			- Vehicle size and load factor
			- Availability of seats
			- Accessibility
			- Travel comfort
			  - noise
			  - vibration
			  - temperature
			  - tidiness

	Route	Route quality (travel of passengers and cargo)	- Distribution and number of stations
			- Timetable:
			  - peak trains,
			  - first-last train
			  - weekend-commuter services
			- Frequency (number of trains per hour)
			- Punctuality/reliability (waiting at stations)
			- Cargo services (reliability)

	Service	Pre-transport and post-transport service quality (added value to service)	- Ticket sales/reservations - Handling - Staff adequacy and competence - Inquiries and general information - Response to complaints
External Quality		Externalities (safety and environment)	- Public service obligations - Safety procedures - Liability regimes - Environment protection (noise, pollution) - Congestion
Dynamic Quality		Investment policy	- Fleet and track renewal rates - Track and stations maintenance - Investment obligations

Taking into account these three characteristics, Box 4.1 summarizes the five most important quality dimensions for the railway industry (vehicle, route, service, social and dynamic quality) together with a number of standard performance measurement instruments for them. The three first dimensions (vehicle, route and service) are related to what is usually named *quality of service*, whereas the other two refer to externalities in a static and dynamic sense.

4.1.1. – Quality of service.

Regulation of the quality of rail transport services in those aspects regarding the vehicle quality, the transport service itself (aboard trains) and the pre-transport and post-transport services has been dealt with in different depth in most countries although, as it was described in Section 2, there is a positive correlation between the extent of the restructuring activity in the rail industry (in terms of private participation and/or separation of infrastructure from services) and the quality regulation requirement imposed to the post-restructuring industry.

In general, countries where the sector is still heavily dependent on government or public agencies have done less to establish a separate quality control framework (as in some countries in Eastern Europe or Asia), whereas in those where private participation has been significant (the case of United Kingdom, for example) a detailed system of quality control has been setup. In all cases, the basic principle governing the design of quality mechanisms is that customer service should be paramount if railways are to maximize profitability and compete with alternative modes of transport. The economic relationship between separate units in a railway enterprise should be structured to ensure that incentives to maximize customer service are preserved (see Swift, 1997a, 1997b).

This is particularly relevant with respect to the separation between infrastructure and operations. Vertical unbundling in railways distances infrastructure management from the end-user customer and it could yield undesirable side effects or contradictions. For example, the density of traffic (trains run per day) that maximizes returns on infrastructure investment is likely to be greater than the level that is optimal from the point of view of operators. This is because at high densities, passenger service is likely to suffer through congestion. Therefore, no matter whether separation is institutional or only financial, mechanisms to compensate infrastructure units from running below optimal capacity must be incorporated in contracts in order to maximize the end-user customer performance as a whole. Since the particular characteristics of the rail industry in each country requires a fine tuning adjustment of any regulatory or contract enforcement mechanism, Box 4.2 proposes a simple scheme that identifies and separates the roles to be assigned to the regulatory agency and the operators (either franchisees or public or private monopolies) with regard to the quality of service regulation.

After its reform and full privatization of services and track provision, the United Kingdom rail system constitutes one of the most practical examples of a detailed quality of service regulation framework in the sense defined by Box 4.2. For example, in the case of passenger transport, the regulatory agency (*Office for Passenger Rail* Franchising, OPRAF) defines what level of service would be tendered for a particular route or corridor. It also sets the minimum level of service for every route in the country (not only timetable specifications, but also journey time, first and last departure times, etc.) If franchisees operate a poorer service than specified then the OPRAF reserves the right to withhold the grant.

Box 4.2. Role assignment in railways quality of service regulation.

Role	Regulator	Operator	Both
Design of adequate quality of service standards	✓	×	×
Level of application of these standards	✓	×	×
Punishments, fines, sanctions,...	✓	×	×
Information to passengers about quality standards	✓	✓	✓
Variables to be controlled	✓	×	×
Inspection and reporting procedures	✓	✓	✓
Responsibility for achieving quality standards	×	✓	×
Risk sharing of service quality fluctuations	×	✓	✓
Technical quality	✓	×	✓

Operators awarded with licenses, *Train Operating Companies* (TOCs), are obliged to include in their timetable certain passenger service requirements set out in the franchise agreement. These are the minimum standards of quality that operators need to achieve to ensure a basic provision of services. However, these requirements do not specify detailed timetables for each route in order to avoid excessively limiting the freedom of the operators, and instead set out parameters within which each company must design its own timetable. Passenger service requirements are set out by route and are largely based on former *British Rail*'s timetable, specifying frequency of trains, stations to be served, maximum journey times, first and last trains, weekend services, through services, and load factors/peak train capacity (for commuter services). Passenger service requirements also include limits on the total number of train cancellations and, where applicable, on the level of capacity which needs to be provided. These limits apply in any 28-day reporting period, with three levels determined: *(i)* a call-in level, at which OPRAF reviews the performance of the operator; *(ii)* a second level, at which the operator is in breach of the franchise agreement, and *(iii)* a third level, which can trigger default under the agreement.

With respect to other quality variables, compliance with load factor requirements, for example, is measured by reference to the ratio of passengers in excess of capacity to the total number of passengers (PIXC). The maximum acceptable level of PIXC is 3% for morning and evening peak together, and 4.5% for either peak considered alone. If extra capacity is needed to meet load factor specifications, the cost is shared between the operator and the OPRAF according to the following criteria: *(i)* up to a certain limit of capacity increase the franchise payment does not change; *(ii)* between the initial limit and a second limit OPRAF bears a share of costs, and finally *(iii)* above the second limit all costs are paid by OPRAF.

In practice, not all quality dimensions defined in Box 4.1 can be incorporated in the same proportions to any service quality mechanism. The British system mainly focuses on the route dimension and is based upon their extensive experience in deregulation. When the

role assignment proposed in Box 4.2 is not considered, or its components cannot be easily separated, several quality regulation failures may arise. The most important one is the failure in defining adequate and independent quality measures. This is the case of several rail concessionaires in Argentina, where the level of vertical integration between the train service providers and the maintenance firms (in the form of subsidiaries or units integrating a larger industrial group) has distorted the incentives to provide the optimal price-quality ratio in favor of more frequent repairs and technical updates.

4.1.2. – Safety and externalities.

Regulation of the quality of service is only one of the two static aspects of quality regulation to be considered in designing a global framework for quality regulation in the rail industry. The *social or external dimension* of quality regulation, which includes all safety and externalities related issues (pollution, congestion, etc.) is the other factor to take into account, and it specifically differs from the level of service dimension of quality regulation in at least four aspects.

The first element is the scope of regulation. Since non-compliance with the social quality standards defined in Box 4.1 may affect users and non-users of the transport services, these standards should always be exogeneously set, and, in the case of the rail industry, often by national or supranational legislation with intermodal implications. This is not always the case in variables such as timetables, load factors or vehicle size, which usually have simple intra-firm consequences. In the European railway industry, for example, three levels of quality regulation can be found: Directive 91/440 determined the overall principles, and the obligation to comply with them was envisaged either in mode-specific regulation (e.g. Railways Act in the United Kingdom) or in legislation that applies to all sectors of the economy (e.g. Health and Safety Act).

The second factor that makes service quality regulation different from social quality regulation in the rail industry relates to the regulatory approach that must be used in the latter. Since the risks associated with accidents or potential environmental damages do not only directly affect the private benefit, but also the social benefit of this transport mode, there is a need for an external regulator or agency to take charge of coordinating safety and reliability procedures at the industry or inter-industry level. This is particularly important with respect to coordination when firms move from a public system to a deregulated one as was described in Section 2. Furthermore, in the railways industry, separation of infrastructure from services, and the introduction of open access, has made it necessary for a rail track controller to ensure safe co-ordination between different operators using the same track or stations.

Again, using as an example the British railway system, their safety regulator is the Health Safety Executive (HSE), which informs and gives advice to the Office of the Rail Regulator (ORR). Operators of railway services, stations and networks must have an accepted *safety case* before the ORR approves their license. A safety case is a complete

resource, control and management plan for delivering safety, defining safety procedures, organizations and systems. The private infrastructure provider, *Railtrack*, is required to have its own safety case, a fundamental component of which is *Railtrack's Safety Management System*, which is a system of standards for controlling operational procedures and technical standards to ensure safety and safe interworking in Railtrack's infrastructure.

The third aspect of particular interest related to safety regulation in the railway industry is the assessment and assignment of risks. Given the inherent difficulties associated to a strict monitoring process, there exist clear incentives for private providers of rail transport services under quality regulation to subordinate their compliance with safety requirements to the previous achievement of financial objectives.

In fact, despite recent tragedies in 1998, railways traditionally have had a high public reputation for safety and this perception has converged with the statistical proof in most countries. Therefore one could conclude that safety levels and management are quite sufficient and no particular safety precaution or in-advance measures should be taken. However the public outcry, negative social effects and adverse public opinion to single catastrophes, together with the persistence of regular fatalities (staff accidents, passengers joining and alighting from trains,...) make it impossible for the regulator to avoid designing measures and policies aimed at diminishing both individual and social risks.

One of these policies relates to the compulsory insurance against third-party liability, since it may correct the above described incentives to excessive risk taking by the operators. In Europe, for example, Directive 95/18 required that operators of train services have to obtain, together with the operating license and path allocations, a safety certificate and insurance. The insurance arrangements in the privatized British railway industry provide another example of scope of liability cover, its basis and conditions for self-insurance. In this case, licenses for the private operators of railway assets (passenger trains, freight trains, stations, and maintenance depots) contain a condition requiring the operator to maintain insurance against third party liabilities regarding licensed activities. The type, cover, level and identity of the insurer need the approval of the regulator who sets guidelines on the minimum insurance requirements that the operators must meet. Operating licensed activities without insurance approved by the regulator is considered a breach of the license.

Finally, the fourth element that makes service quality regulation different from social quality regulation is related to externality issues and, in particular to those connected with the environment (engine pollution, noise, transport of hazardous goods, etc.) Again, in this case, social quality regulation should be concerned with both internal and external factors of the rail operators, and have several differences and similitudes with other transport modes.

For example, air pollution is usually one of the most regulated areas with regard to quality regulation in road and air transport modes. In the rail industry, however, air

pollution is not a critical issue, with few notable and restricted exceptions in certain countries or routes. Noise pollution, either in suburban neighborhoods, areas close to stations, depots or particularly delicate countryside ecosystems has attracted a larger attention both from the public and the regulators. Most countries, therefore, incorporate in their regulation the design and specification of measures to reduce noise produced by rolling stock and stationary sources (fans, compressors, and generators) and shunting noise.

Two final related issues with regard to environmental regulation are the measurement, analysis and prediction of emissions of chemical substances resulting from the presence of railway lines (heavy metals, lubricants, dust, etc.), and the assessment of the risks to the safety of local residents as a result of rail-related activities (transport of dangerous goods, explosions, etc.) In these cases, as in many other aspects, most countries subordinate their social quality standards and the role of their regulators to the overall technical principles emanating from supranational organisms or professional associations. Private and public rail transport operators are obliged to comply with the environmental standards set at the national and supranational level. In Europe, for example, there are EC Directives on air pollution from vehicles specifying the environmental standards authorized to vehicle engines and fuel qualities, and these apply to both vehicles (wagons, locomotives) and transport operations.

4.1.3. – Dynamic quality: investments

According to Box 4.1, there exists a third dimension in the quality regulation of the rail industry. Since the regulatory process is by itself a dynamic relationship between, on one side, the regulator and, on the other, the regulated transport providers and the firms and their passengers, the dynamic links of such a relationship must be taken into account when certain quality standards are to be controlled. In particular, the investment policy of the railroads is the most important dynamic element to be considered in the design of concession contracts and, in particular, with respect to the implications that these investments will have on the future performance of the firms. A complete quality regulation regarding the investment policy must start by defining which part decides the investment objectives, in terms of fleet and track renewal rates, track and station maintenance, or future investment obligations, for example. When the regulator assumes this role, it must also set up adequate mechanisms to monitor the progress of the investment stages, and provide the incentives to avoid stopping projects before their conclusion.

When the regulated rail transport provider is in charge of its investment policy (with respect to the renewal or maintenance of its fleet, for example) a quality control should also be imposed to avoid, for example, inadequate planning or excessive unnecessary repairs as a means of earning extra revenues from subsidiary companies. In some countries, most clearly in the United Kingdom, this control has been exerted by isolating non-commercial investments and non-commercial aspects of overall investment planning, and making them the subject of specific grants from public funds.

In all cases, one of the most controversial issues to be carefully discussed in the concession contract is the relationship between this investment and the prices to be set to

recover it, since *ex ante* prices are decisive in determining the extent and mix of investment in new rail infrastructure. Uneconomic investment decisions have, historically, been imposed by governments on railways, and in most countries this has been the main cause of accumulated debt. In fact, as described in Section 2, insulating railway operators from such debts has been a central aim of the railway reform and restructuring processes carried out around the world. Thus, in principle, the simplest decision rule, according to standard economic principles, is to proceed with projects whose net present value, calculated according to a suitable discount rate, is positive. In theory, the most obvious discount rate to use in public sector projects is the interest rate on long-term government bonds.

In practice, however, this bond rate may not be appropriate in several circumstances and governments which choose a discount rate lower than this rate invariably find that they cannot afford to proceed with all the projects with positive net present value.[20] This means that, in certain cases, it is more useful to specify a hurdle rate (that determines whether the project will in fact be implemented) and then test it against the project's internal rate of return. Therefore, on purely economic efficiency grounds, if this selective process is strictly conducted and calculations are correct, only projects with non-generating losses on new investments should proceed in order to avoid later problems with cost recovery in regard to infrastructure pricing. However, on social grounds, few rail investments would pass this strict cost-benefit analysis, and subsidies to pay for the fixed costs may be required at the investment point.

With regard to investment financing, the more important aspect to consider from the point of view of quality regulation in the concession contract design, is the monitoring of the financial health of the operator so as to prevent possible cheating incentives (for example, lowering the quality of building materials in tracks, signaling mechanisms or stations). In principle, there are no specific criteria for rail investments according to which the regulator should impose particular rules with respect to the firm's capital structure. When the size of the investment is large enough, it is the private concessionaire who will seek the adequate mix between debt and equity that enables her to carry out the project. Only if there is the participation of the government or some public agency in the new investment, should the concession contract regulate the conditions and terms under which the assets transfer (if any) takes place.

Alternatively, within a context in which the main (sometimes, unique) source of funding for the infrastructure provider is the revenue from track charges, different criteria are to be used to reflect two basic circumstances: where the aim is to maintain existing standards, or to provide increased capacity and better quality of service. In the former case, the regulated access charge is set to provide the cost of renewal as seen by the regulator. The incentive for the infrastructure provider to undertake the investment is provided by the performance regime. Where expansion or improvement of rail capacity is required, the track manager is expected to finance such investments by means of increased revenues, so that train operators and infrastructure providers should share both the risks and benefits of an improved infrastructure.

[20] See for example Layard and Glaister (1994), for a description of standard cost-benefit analysis procedures.

Finally, as mentioned in previous sections, the clarity and simplicity of the negotiation and re-negotiation rules constitute a relevant element in the dynamic relationship established between regulator and operators. A common situation, for example in Argentina, is that once licenses have been awarded, rail operators use fake or real (but possibly not required) quality investments to improve their position and demand changes in their license conditions.

4.2. – Instruments for quality control in the rail industry.

Once the quality objectives, in its service, social and dynamic quality dimensions, have been setup, the following step in any process of devising a system for regulating quality in the railway industry concerns the design of the instruments for control. In principle, there are three alternative mechanisms for regulating quality in the railways industry.

First, the firm could be required simply to publish measures of quality every pre-defined period and to report this information to the regulator. This information can also be made public to inform consumers and/or actual or potential rivals about the operator's current performance. As in any other type of regulation process, access to public information is a very delicate issue since it can both serve as a disciplinary device for the rail provider and a strategic instrument to undermine or strengthen the ability of the firm to survive in the market.

A second mechanism to control quality consists of including a direct and explicit measure of quality in the price control mechanism. For example, when subject to rate of return regulation, a rail service provider may be obliged to calculate its asset base according to certain average values and/or require authorization to carry out certain technological improvements in order to avoid overinvestment to make use of the Averch-Johnson effect. Similarly, under price-cap restrictions, the basket of products whose average price increase is controlled by the regulator can be suitability defined in order to avoid the changes in quality (and consequently, cost reductions) that could be used by the regulated firm to increase its profits, even if the same price caps are maintained.

The third mechanism that can be used to control quality is the setup of customer compensation schemes, through which grants or payments are awarded to those affected by non-compliance with quality standards. In practice, however, these mechanisms only work if quality failures can be easily verified. This requires a detailed regulation not only of the quality standards, but also of the monitoring rules and the guarantees for the regulator and the regulated that the inspection process will be transparent and objective. Moreover, if the compensation is distributed to the consumers, either by the firm directly or through an intermediary body, sharing rules must be also defined. The practical difficulties associated with this quality control mechanism have made it common in many countries to specify instead the minimum quality standards for certain parameters of its rail industry, backing them with explicit legal sanctions that may include fines or the revocation or withdrawal of the license to operate.

Box 4.3. Instruments for quality control in the rail industry.

Regulation stage	Instrument	Additional characteristics
Stage I **Before entering the market**	- Pre-tender qualification requirements	- Experience - Financial strength - Technical ability
	- Specification of service characteristics in licenses	- Routes and frequencies - Timetables, - Vehicle capacities and load factor - Punctuality and reliability
	- Specification of financing rules and investment plans	- Investment plans - Fleet and track renewal rates
Stage II **During market operation**	- Quality in price-control mechanisms	- Rate of return regulation vs. Price-cap regulation
	- Information revelation obligations	- Control of access to critical information
	- Audit processes	- Internal and/or external
	- Company reporting	- Frequency - Format
	- Regulator's direct monitoring	- Setup of monitoring mechanisms and rules
	- Technological control	- Tacograph readings, use of electronic controls.
Stage III **After market operation**	- Incentive payments	- Customer compensation schemes
	- Penalties	- Fines for underperformance
	- Enforcement and binding rules	- Contract withdrawal as a last resource

Finding the adequate mix of these control mechanisms often becomes the most difficult task in the design of the quality regulation process. The approach followed by most countries in their railway restructuring process can be outlined in Box 4.3, where the most important instruments are summarized.

In conclusion, the quality regulation process comprises three stages. First, before entering the market (Stage I), the aim should be to anticipate and minimize future conflicts between the regulator and the concessionaire of the rail service.[21] Licenses must specify the expected characteristics of the service in terms of, for example, routes and frequencies of trains or timetables. For passenger services, particularly in the case of urban or suburban trains, vehicle capacities and punctuality can also be set. Finally, in order not to forget the dynamic dimension of quality described above, Stage I must also specify investment plans and financing rules. Afterwards, during market operation (Stage II), instruments for quality control in the rail industry should mostly be related to the direct monitoring of the firm's performance. Thus, this is the time to introduce incentives toward quality in price-mechanisms, to design the information revelation obligations of the firm, and the auditing (external or internal) processes to be carried out. In most cases, the use of instrument of technical control (such as tacographs or track electronic controls) complements the standard instruments. Finally, after the transport activity has already occurred (Stage III), compensations or punishments can be implemented. Several schemes, as described above, can be used. In any case, both penalties and incentives must be graded according to the expected future evolution of the relationship, since severe fines or large subsidies may alter the standard behavior of the operator in the market.

5. – THE ROLE OF RAIL INFRASTRUCTURES.

The separation of infrastructure from operational services in railways constitutes a relevant feature in this sector, conditioning the concessioning process in many countries. The vertical unbundling process, aimed at solving the natural monopoly issue described in Section 1, cannot only promote greater allocative efficiency, but also encourage some other questions of relevance for regulation. These questions become particularly important when the degree of private participation in rail infrastructure management is relevant. In this case, the economic regulation of infrastructure should be governed by the adequate combination of three standard principles: fair access to the infrastructure, cost recovery, and efficient access pricing.

5.1. – Access to rail infrastructures.

Regulation of rail infrastructure does not only include simple pricing principles, but also access rights and long-term development provisions. The form in which these subjects have been addressed varies among countries. Most governments have opted to retain infrastructure in public ownership with the creation of a State agency to manage it (the first example is Sweden's *Banverket*) with respect to private train operators (as in Argentina); others (such as France and Germany) have established independent State-owned enterprises to manage railtracks. Only the United Kingdom has privatized both infrastructure and operations.

[21] To achieve this, pre-tender qualification requirements can be used in order to ensure a minimum level of technical and practical expertise and financial solvency, as described in previous section.

Whatever the case, either in public or private hands, the infrastructure regulation of the railways industry should first start by including minimum investment requirements to counterbalance short-term myopia where infrastructure has been privatized, in order to ensure that key investments are not subordinated to the private goals of augmenting annual dividends or defending the firm from potential hostile takeovers.

Secondly, the regulation must tackle 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 countries in Asia). In the European Union, for example, Directive 91/440 directs each member State to grant international access and transit rights to international groupings in which 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 has advocated that these provisions should be extended to all freight and international passenger services. Most countries simply charge their (still monopolistic) train operators for the use of the (also public) rail infrastructure.

In the wholly privatized structure of the United Kingdom, open access has been limited for passenger services by a number of provisions regarding moderating excessive competition. These provisions were initially designed to provide protection to rail franchisees both from new entrants and from each other, and it is anticipated that they will be gradually reduced over time. In other countries, access rights are also clearly specified in the contract, as mentioned in Section 3. For example, in Argentina the concessionaires have exclusive use of the tracks, but must grant access to passenger operations in return for a compensatory track fee. In the Abidjan-Ouagadougou rail link between Côte d'Ivoire and Burkina Faso, after the initial seven-year exclusivity period the current operator will have to grant track access, for an agreed fee to any third party carrier approved by the regulator (Mitchell and Budin, 1997). In certain large cities, like Mexico, D.F. or Buenos Aires, the operators share a common network under a unique transport authority.

The final aspect regarding access rights to rail infrastructures lies in the removal of existing or potential barriers to entry that might distort competition by favoring certain competitors against others. These barriers should also include technical requirements (related, for example, to incompatible rolling stocks and tracks) and safety standards (in terms of the common minimum level to be attained by all operators). In summary, the general rule should be to promote open access as widely as possible, once the separation between the natural monopoly infrastructure and the train operations has been effectively achieved. However, this process cannot be made independent of a detailed analysis of the infrastructure costs and the prices charged for covering them.

5.2. – The problem of rail infrastructure costs.

As described in Section 1, rail infrastructure provision and management are characterized by a high ratio of fixed to marginal costs, the existence of avoidable costs and non-avoidable or

common costs. Avoidable costs are uniquely associated with a particular output: were this output not produced, no cost would be incurred. The use of this guiding principle relates to the idea of cost recovery for particular outputs. Avoidable costs may thus be considered as a floor to regulated prices (if any), since charging less than the avoidable cost is equivalent to operating at an economic loss. This makes standard pricing rules inoperable in this sector, since first best or efficient principles of marginal cost pricing may derive in large deficits that jeopardize the survival of the firm in the long run. Three particular problems then arise with respect to the allocation of the rail infrastructure costs: cross-subsidization issues, cost-recovery problems, and the possibility of setting inefficient prices for infrastructures (Talley, 1988).

The existence of cross-subsidization problems in pricing rail services or infrastructures produced in the presence of common costs, can be illustrated in the case of a profit-regulated railroad which interconnects two large cities and also provides rail service to a smaller town located on the route between the two cities. The fares charged for passage from the small town are sufficient to generate revenues in excess of the additional costs of serving it, such as ticketing and station costs, but not sufficient to cover an equal or proportionate (however defined) share of the common costs, such as trackage, signaling, and trainyard costs. The issue is how to allocate common costs among customers and services. In many cases, cost sub-additivity also, and then efficiency requires joint production and allocation of fixed costs among all services, without cross-subsidization (accounting for externalities whenever present).

Cross subsidization is not only an equity problem for rail services, as in this example, but also a particularly relevant issue for efficient pricing in infrastructures like railbeds, signals or stations. The standard procedure is the so-called fully distributed costs method, under which common costs are allocated on the basis of some common measure of utilization, such as gross tons-km, or other measure of relative output or gross revenue. Alternatively, common costs can also be allocated in proportion to costs that can be directly assigned to the various services (Braeutigam, 1980). The arbitrary nature of fully distributed cost methods and its lack of conceptual foundations have been criticized, but they remain a useful measure for recovering common costs.

However, the treatment of the cross-subsidization problem should not be based on excessively rigid criteria, particularly for developing countries with few alternative finance mechanisms. The analysis should be made on a case-by-case basis, since, for example, stand-alone cost test do not apply if railroads are not allowed to abandon unremunerative facilities or services (Kessides and Willig, 1995). Where that freedom is denied, a railroad cannot earn adequate revenues if its rates on potentially remunerative activities are constrained by stand-alone cost ceilings.

With regard to the cost recovery principle, this should be a central issue in the design of any rail infrastructure pricing procedure. The theoretical and political debate has focused on two options. On one hand, many public firms still advocate the use of efficient price mechanisms (in the sense described in Section 3) and propose marginal cost rules with the simultaneous use of public subsidies to cover fixed costs.

Alternatively, a growing literature patronizes the use of full cost recovery prices, including price discrimination, multiple part tariffs or cross-subsidization schemes, if needed. Although it is reckoned that it might yield inefficient outcomes with regard to the theoretical efficiency principles, they constitute the second best available alternative in most cases.

Similarly, with respect to access pricing to a rail network, it is clear that it should be based on marginal cost pricing rules in a first best world. In practice, however, the achievement of this objective is difficult due to at least three reasons: the above described cost structure of the rail network, which cannot always be recovered with simple price rules; the asymmetric information problem faced by the regulator with respect to these costs; and the subsidy level that can be sustainable in the long-run.

Many econometric studies have shown that in the case of the rail industry, the marginal cost of still vertically integrated railways lies in the range of 60%-70% of average cost; where rail services are separated from infrastructure, the marginal social cost of rail infrastructure alone often is well below the 60%-70% range. (Friedlander *et al.*, 1993). Price discrimination, if feasible and politically acceptable, might succeed in raising cost recovery to around 60% of total cost without driving demand off the market. Thus, full cost recovery would require a further price-mark-up of more than 60% above the efficient price. Alternative proposals, in terms of the so-called Ramsey pricing principle, have been defended for infrastructures with high fixed costs and low marginal costs.[22] However, they rarely work in practice, since they arouse consumers' suspicions of unfair treatment and undue discrimination. Moreover, under Ramsey pricing rules all unattributable fixed and common costs are apportioned on the basis of the services' demand characteristics.

In the current debate, a reasonable conclusion seems to advocate the adequate balance between the cost recovery issue and the efficient pricing rules, giving some preferential treatment to each one according to the case. In general, the issue remains unsolved and depends on how different countries have faced their access pricing problem. Whether a country's government is willing to assume these differences or not is, in most cases, a political question. In many cases, the ultimate challenge is how to price access to rail infrastructure in a transparent, efficient and non-discriminatory way. In Europe, for example, Directive 95/19 requires infrastructure managers to balance revenues with expenditures. In countries where revenues from operations and compensation from government for public service obligations are insufficient to provide a surplus for depreciation and investment, railways will be dependent on the State to fund or guarantee repayment on loans for investment. This continues to be the case in many of the countries of central and Eastern Europe.

5.3. – The access pricing problem.

The development of tariffs for accessing rail infrastructure varies greatly among different

[22] Ramsey pricing refers to charging larger prices above unit costs to more inelastic market segments. When infrastructure and services are separated, their use becomes more complicated and still is not clearly solved, since different demands for services – as well as for tracks – must be estimated.

countries according to the stage of their railway restructuring process. Some countries have already identified procedures for setting fees, and a number of them have laid down precise rules for the structure and level of fees. In others, business unit or infrastructure companies (either in public or private hands) are responsible for setting charges. In fact, access charges are mostly relevant in countries where traditional railroads have been vertically unbundled by the separation of the potentially competitive area of service operations from the naturally monopolistic area of infrastructure management.

Apart from the already discussed problem of cost recovery, access pricing may create a market structure problem regarding its effects on competition and barriers to entry. This problem arises in network industries where a single vertically integrated dominant firm (either private or public) controls the supply of a key input (in this case, railway tracks) to its competitors. It is obvious that in these cases, there are incentives for the firm to set prices high to raise rivals' costs, but it could also be the case in which the regulator sets access prices too low in order to favor the entrants.

In practice, these potentially distortionary effects on access prices can be determined in several ways, depending on the discretion allowed to the integrated firm. First, when the infrastructure is still publicly owned or managed, the regulator can determine the price as an integrating part of the access terms defined in a contract with one of several private train operators. Secondly, the regulator may allow the firm to choose from a menu of alternative regulative schemes, usually based on incentive-based price regulation mechanisms (in order to favor the firm achieving higher levels of efficiency). Thirdly, the firm may have discretion over aspects of access pricing subject to some overall regulatory constraint, and finally, the firm may have full discretion over the price and is only restricted by the provisions of the antitrust law of the country.

In any of these cases, there are two main approaches to setting-up access pricing when the principles of cost recovery plus the normal rate of return are required. First, there are countries that use *cost-related charges*, which constitutes the current dominant paradigm for setting access charges. These charges are based on the optimal first-best principle of pricing according to marginal cost, which is considered the forward-looking long-run incremental cost.

This principle becomes more complex the higher the proportion of common costs, and is based on the so-called *efficient-component rule*, which determines that the optimal access charge should be equal to the direct cost of providing access plus the opportunity cost of providing access (given by the reduction in the dominant firm's profit). To compute these costs the regulator will also have to consider economic depreciation (physical depreciation plus technological progress), and forecast the likely future usage of the elements.

The first problem to be solved is that of the actual value of capital assets: nominal value versus potential to generate cash. While the latter is clearly a function of selected methods of privatization and regulation, and the extent of competition envisaged to bid for the right of operating concessioned infrastructure services, the former option is more likely to reflect a past

situation that domestic reforms are trying to overcome.

The second method of setting access prices consists in developing *usage-related* charges. Once avoidable costs are covered in increasing prices inversely related to demand elasticity. Another option (a less controversial approach, particularly if it is optional) is the use of a two part tariff to avoid train operators cutting services to save charges even when there is no cost saving in the network.

The case of the British infrastructure provider, *Railtrack*, constitutes a well-studied example of the functioning of access prices in practice. The main targets in the constitution and privatization of this firm were set to obtain a better organization of transport services, reduced costs and a higher efficiency. In a context where operating companies have also been franchised, *Railtrack* manages the infrastructure (that is track, signaling systems, electric power supply and stations) and is responsible for its maintenance, new investments and train operations (timetables, coordination, etc.). It also sells access to infrastructure to the (passenger) operating companies and to the freight operators.

Railtrack owns the rail network and set track charges that have to be agreed upon with the rail regulator under the criteria openly published through a number of regulatory policy statements. The price control system operates through a simple RPI-X formula, which is revised every five years and remains fixed until the next revision. For example, in January 1995 the regulator announced the price controls that would apply to franchised passenger services from April 1995 to April 2001.

The structure of *Railtrack*'s access charges for franchised passenger services is based on the principle of usage-related charges and, in particular, are made of multiple part tariffs, with at least four elements.[23] First, track usage charges, which tend to reflect short run effects on maintenance and the renewal costs of running trains of different types for different distances. Second, traction current charges, which recover the costs of electric current, varying geographically and temporally and reflecting distance covered and type of vehicle. Third, the long run incremental cost, which indicates the long run costs imposed on *Railtrack* in delivering the total access rights of a train operator; and finally, the apportionment of common costs, which is a remainder of the fixed charge designed to recover the rest of *Railtrack*'s costs at the sub-zonal, zonal or national level. This is apportioned among train operators on the basis of budgeted passenger vehicle miles for sub-zonal costs and budgeted passenger revenue for zonal and national costs. Whereas the first two elements amount on average to only some 9% of total track access charges, and given the current structure of charges, only these elements vary directly with use. The remaining 91% of the aggregate charge is in the form of a fixed charge, which does not vary with the number or type of train runs, or with passenger revenue.

In the case of freight services, access prices are more flexible. The rail regulator has simply established several principles to be considered by*Railtrack* in its relationship with private operators. First, the prices must cover the avoidable costs incurred by *Railtrack* as a direct result

[23] See Dodgson (1994) and ORR (1997) for a detailed description of the British system.

of carrying that particular freight flow; second, the prices must be lower than the stand-alone cost, which would be incurred by a notional efficient competitor; third, no undue discriminatory charges are possible; and finally, the structure of charges should reflect the value to users of access to the rail network, and enable *Railtrack* recover its total cost

As opposed to the British case, the setting of access charges in other European countries is still underdeveloped. In 1995, the European Union passed two directives concerning the application of Directive 91/440 on the separation between infrastructure management from transport operations. Directive 95/18 regulated the licensing of railway undertakings, whereas Directive 95/19 established several general principles on the allocation of railway infrastructure capacity and the charging of infrastructure fees. These principles were designed to ensure an optimum and non-discriminatory use of infrastructure and guarantee an access charging policy according to EC rules, but Member States received these Directives with various degrees of enthusiasm. The objective of most governments that have set rules for infrastructure fees is to cover costs, differentiating fees to reflect such factors as type of service, wear on track, distance of run, routing, etc.

In France, for example, several principles were introduced to give access to railway infrastructure for licensed international groupings of transport services and operators of combined transport, but present arrangements seem keener to promote conventional rail international groupings rather than new entrants in the rail market. With centrally planned timetables, the only domestic operator pays a fixed amount to the (also public) infrastructure manager. User fees are fixed taking into account a wide set of criteria, including infrastructure costs, the transport market situation and characteristics of supply and demand, imperatives based on optimized use of the national rail network, and the standardization of conditions for intermodal competition.

Similarly, in Germany, the Federal Government owns the track infrastructure and is responsible for its preservation and for securing a certain level of service in public transport by means of the *Deutsche Bahn* (DB), an independent joint stock holding whose sole shareholder is the State. The infrastructure division of DB bears the costs for operating and maintaining rail infrastructure, and it is also in charge of stations, sale of tickets, passenger attention, etc. It is also responsible for setting charges for the use of the track, which are supposed to cover all infrastructure costs, including investment. These charges are based on prices per train-km on the different line sectors, which results in a considerable number of different fee combinations (Häfner, 1996).

5.4. – Coordination and intermodal competition.

A final relevant issue when considering rail infrastructures and its pricing is the topic of inter-modal competition. As it was mentioned in Section 1, modal choices could be heavily distorted due to different cost coverage ratios and the use of different bases for cost imputation.[24] To solve this problem an integrated, multi-modal, approach should be followed. Basic principles will have

[24] In fact, one of the reasons mentioned in Section 2 to explain the decline of the rail industry was the fact that road transport did not internalize its social costs (in terms of pollution or safety, for example).

to apply to all transport operators, irrespective of the mode in which they operate. For example, in countries like Argentina or 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 find themselves unduly disadvantaged by competitors who enjoy public aids or indirectly benefit from huge externalities.

In the case of natural monopoly infrastructures, the principles envisaged to avoid these distortionary effects should be solidified in the coordination between existing networks (particularly, in dense rail areas) and the setup of mechanisms that facilitate inter-operability and international links. On this score, neither the most advanced infrastructure regulations, such as the Swedish or the British systems offer much help, since their were conceived for a single country environment. In other countries, such as Argentina before the restructuring process, railways were directed to solve national transport problems (in terms of offering underpriced 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 slots assignment on more congested corridors.

In summary, the main conclusion of this section is that the control of track access, pricing and development of infrastructure is an inevitable part of any rail industry where vertical unbundling and a notable degree of private participation in train services provision has been achieved. In these cases, the ultimate goal of regulation should be to ensure that infrastructure access and its pricing promote an efficient structure of production, use and consumption of the transport services, while allowing network providers to make a sufficient return.

6. – PERFORMANCE INDICATORS IN THE RAIL INDUSTRY.

6.1. – The use of performance indicators in the rail industry.

Performance indicators in the rail industry are used to monitor the behavior of one or more regulated firms in order to evaluate the effectiveness of the regulatory measures they are subjected to.[25] The main advantage of these indicators or indices is that they provide a periodical assessment and control of the firm's activity, and a system of continuously updating information, that is simple, quickly obtained, and has a relative low administrative cost for the regulator.

Their most important disadvantage is that their use is only valid when comparisons, either between different firms or the same firm over time, are constructed on a similar basis. For example, when inter-firm comparisons are to be made, the companies must belong to countries with similar characteristics (e.g. in the participation of transport in the economy as a whole, the degree of economic development, or the regulatory framework, etc.), whereas if intra-firm comparisons are intended, the indicators must account for external and internal changes produced during each period (e.g. new management, changes in demand, etc.)

[25] For example, quality indicators, in the sense defined in Section 5, can be established in a contract and reviewed regularly to confirm that the terms of the license are being fulfilled.

Comparisons across companies usually provide persuasive and very interesting results that can serve the regulator in setting their objectives and designing future license contracts. However, extreme care should be used in drawing normative conclusions from these results. What constitutes a benchmark of desirable practice for some objectives may be different among companies. For example, countries with very liberalized frameworks in their rail industry (as the United States, for example) could set levels of desirable practice for productivity indicators (or quality of service) that clearly differ from the levels in other more regulated frameworks (as Europe).

Similarly, simple indicators should be carefully interpreted over time to avoid incurring contradictions and inappropriate measurements. For example, when assessing railway output, the number of train-km might be a relatively high number, while the number of passenger-km or tons-km may be relatively low (if the firm specializes in one type of traffic). Given this conflict, the notion of overall performance can be ambiguous and the most practical solution is a joint interpretation of the indicators and the objectives that they serve. For example, a service quality objective, such as the number of trains per hour, may conflict with financial objectives, reflected in the high cost recovery rate, or with objectives based on the maintenance of low prices.

Thompson and Fraser (1996) point out that both monetary and productivity variables should be carefully defined when attempting inter-firm comparisons. Fares, wages, outputs and inputs widely vary among countries for a large number of reasons that are sometimes not related to the firm's operations, but to measurement or statistical errors. For example, average passenger fares are based on the overall mix of passenger classes (with different prices each). Tariffs are often higher per passenger-km for short trips than for long ones, and they must also depend on the existence of government subsidies or artificial compensations. Similarly, common mistakes in the case of freight tariffs include not taking into account the different commodity mix, size of shipment or length of haul. This latter effect, which also affects passenger traffic, is particularly relevant, since some costs (ticketing, billing, or station maintenance, for example) are fixed with respect to the length of the trip but vary with size or distance.

These difficulties are even increased in the case of measuring productivity, since a simple comparison among partial measurements of output cannot capture the complexity of relationships or the variety of productive structures that take place within a rail operator. For example, a commonly used productivity indicator, such as the number of passengers-km or tons-km per employee,[26] depends on such diverse factors (e.g., regulatory environment, structure of the labor market, availability and quality of infrastructure, alternative transport modes, etc.), that it could be seriously misleading if interpreted without care.

To elude these sorts of problems, the construction of performance indicators should avoid excessively simple data management, and use statistical techniques that take into account the differences in the relative environments faced by each company. Oum and Yu (1994), for example, estimated different efficiency levels for a sample of railway companies of the OECD, by introducing internal factors (such as the characteristics of outputs) and external factors

[26] The term *employee* can also refer to terminal staff, administrative staff, train crew or maintenance staff. Similarly, capital can be disaggregated into trains, wagons, terminals, platforms, routes, etc.

(difference in the legal and regulatory framework between companies).

6.2. – Main types of indicators in the railways industry.

Despite these difficulties, a large number of indicators are commonly used nowadays to monitor the performance of firms within the rail industry around the world. The definition of each particular indicator depends on the objective it must achieve and its informative value.

Box 6.1. Contextual indicators in the rail industry.

Type	Examples
Overall economic activity	GDP GDP per capita Urbanization degree Industry structure Energy costs Private cost of capital
Transportation sector importance	Participation of transport in GDP Intermodal market share (passengers and freight)
Overall rail sector indicators	• Output Passenger train-km Freight train-km Passenger-km Ton-km • Revenues Passenger revenue Freight revenue • Network indicators Length of line Length of track Electrified track (%) Route-km/km^2 • Density and service Train route-km per capita Train-km per route-km Average size of shipment Average length of haul • Organization of the industry Regulatory agencies (number) Separation of infrastructure and services (type) Access and entry system (type)
Regulatory and institutional system	State involvement in economy (in % of GDP) Tax and Judiciary system (corruption index)

Since there are several external factors, which vary widely from country to country or from firm to firm, that substantially influence their comparison, a first type of indicators, *contextual indicators*, are required to assist this comparative analysis and to define desirable levels of performance. These indicators include the social and economic characteristics of the railway sector as well as other elements associated with the economy as a whole, and are mainly

directed to the regulator. They control for the exogenous factors embodied in both inter-firm and intra-firm comparisons. Box 6.1 presents several examples that could be obtained relatively easily from international statistical sources.[27]

Box 6.2. Management indicators in the rail industry.

Type	Examples
Commercial	• Prices Average passenger fare (revenues per passenger-km) Average freight price (revenues per ton-km) • Quality of service[28] Average train-speed (in passengers and freight) Delayed arrivals or departures (as % of scheduled) % of lost or damaged freight Average passenger load factor Traffic density (trains per hour) • Pollution and safety Rate of fuel usage (per train-km) Level of noise Level of emission of pollutants Number of accidents or incidents
Operational	• Labor productivity Passengers-km per employee Tons-km per employee Passenger train-km per employee Freight trains-km per employee Total trains-km per employee • Capital productivity Number and kms. traveled by locomotives Locomotive availability (in %) Tons-km per wagon-km Wagons-km per wagon Tons-km per wagon
Financial	• Efficiency Costs per employee. Costs per unit of capital Unit cost (per passenger-km, ton-km, train-km) • Profits Revenues/costs Subsidies

Simultaneously, there are many indicators (particularly, those about prices and quality of service) that are both informative to transport users and constitute inputs for the regulator's control tasks. Jointly with the contextual indicators, these *management indicators* provide the necessary instruments to judge the management and the behavior of the company, and can be grouped at three different levels, as summarized in the examples in Box 6.2. These boxes only provide indicative examples of possible performance indicators for the most relevant categories

[27] In particular, the International Union of Railways (UIC) publishes a yearly summary of the main statistics of its affiliated railways, although not all of them are always available for all railroads.
[28] See also Section 5.

in which a rail regulator may be interested. In the context of concession contracts described in previous sections of this paper, where a number of private companies are potentially franchise operators in the rail industry, the regulatory agency should define those indicators that best suit its information needs.

Some final practical rules that could be helpful in this process are the following ones: *(i)* each indicator should have at least a function or objective, *(ii)* the relationship between each indicator and its objective must be clear and direct, although *(iii) m*ultiple objectives can be addressed by multiple indicators (jointly interpreted); finally, in order to assure the utility of the indicators, *(iv)* appropriate data must be provided, and *(v)* the management of indicators' information should be part of the regulatory process.

Despite the difficulties mentioned, price indicators can constitute for the regulator a mechanism of control over the activities of the operators. This control may be established not only in terms of the comparison between companies of similar characteristics, but through monitoring over a period. In any event, the regulator must ensure that any variation in price corresponds to a proportionate variation in companies' costs or level of efficiency. The operational and efficiency indices are therefore instruments which help the tasks of the regulator. Improvements in companies' productivity and efficiency levels combined with increases in price levels are clear signs of abuse of market power on the part of railway operators.

Indicators of service quality were developed in Section 5, and they should serve, as price indices, to establish evaluations of different companies, as well as dynamic or time evaluations. These measurements should be analyzed together with price indices, because of the possibility of finding different feasible combinations of price and service quality. For example, a high number of trains per hour, i.e. a high traffic density, could only be financed by means of high prices.

The simultaneous implantation of systems of control for prices and service quality may limit the management of the companies and reduce their operativity. Placing the emphasis on the control of prices or of service quality depends fundamentally on whether it i prefers to offer the service at the lowest price possible, or to offer the service with certain standards of quality. All these indicators allow the regulator to monitor the activities of the operators as defined in Phase II of Box 4.3. Unjustified or a systematic breach of the standards of quality (insufficient number of trains per hour, problems of punctuality and reliability of the service, very high indices of load factor, etc) should be accompanied by an appropriate system of penalties, as described in Section 5.

6.3. – Best practices in the rail industry.

Taking into account the comparison caveats made above, the final part of this section compares some of the most relevant performance indicators for the rail industry with those from the companies that obtain the best results, or at least with indices that one could consider desirable. This procedure, commonly used by many governments around the world as a yardstick mechanism, is exemplified in Box 6.3. The last two columns show the values of the best practice levels for a sample of European, Australian and American rail companies, and the values considered desirable according to a World Bank study by Gannon and Shalizi (1995) and the

Australian Bureau of Industry Economics (BIE). [29]

Box 6.3. Best practices in railways management indicators. Examples.

Type	Indicator (example)	Best practice	Desirable
Commercial indicators	**Revenues and prices**		
	Passenger revenue/ passenger-km (in US$)	0.036	0.04
	Freight revenue/ ton-km (in US$)	0.019	0.03
	Freight to passenger tariff ratio (%)	-	-
	Commercial services: general		
	Average train speed (in km/h)	-	60-90
	Arrivals with small delays (among 10-15 min.) (in %)	96	90-95
	Ratio of lost plus damaged freight (in %)	1	1
	Commercial services: passengers		
	Number of passengers per train.	197.5	-
	Passengers-km. per route-km (total)	5237	-
	(in thousands per km)	(136)	-
	Commercial services: freight		
	Number of tons per train	604.13	-
	Tons-km per route-km (total)	2819.19	>2000
	(in thousands per km)	(352)	-
Operational indicators	**Labor productivity**		
	Passengers-km per employee	-	-
	Tons-km per employee (in thousands)	11000	>750
	Pass. train-km per employee.	-	-
	Freight train-km per employee.	-	-
	Total train-km. per employee	4434.84	-
	Capital productivity		
	Availability of locomotives (in %)	-	>80%
	Ton-km per wagon	914.28	-
	Wagon-km. per wagon	-	-
	Freight and passenger wagons availability (in %)	-	>90%
Financial indicators	**Cost coverage**		
	Costs covered with total revenue (in %)	-	>100
	Costs covered with typical revenue (in %)	-	>80
	Cost reduction required to reach break-even (in %)	-	<0

Source: Gannon and Shalizi (1995) and BIE (1995). Desirable values are only approximate and should be taken as general references that might vary across countries and regions.

One of the most useful insights that can be provided by examples like those in Box 6.3 is to make it clear that setting desirable values for indicators is a difficult task. Extreme care should be put into making exclusionary comparisons. For example, according to figures, the unit revenue ratio has a desirable value lower than 4 US$ cents (in passenger traffic) or 3US$ cents in freight. Lower values, as those achieved in several European countries (3.6 and 1.9, respectively) could be indicative of lower prices, or little fare collecting efficiency. In any case, the regulation in force in each country will notably affect the levels of prices charged by each company.

With regard to average train speed, this measurement should distinguish between passenger and freight transport, and among their different categories (urban, regional, long

[29] The BIE publishes every year a benchmarking report (see BIE, 1995, for example) in which their main utilities (including rail transport) are compared worldwide.

distance, international, etc). The desirable indicator estimates an average speed of 60-100 Km/h, but in each country it should depend on the type of traffic, on the social and economic level of each country, and the relative importance of the railways in its development.

Similarly, the measurement of the average passenger load, in terms of the number of passengers per train, is influenced by the different regulatory policies in force, and by other variables such as the size of vehicles and the type of journey. In the case of Europe, the highest level for the year 1994 was attained by the Italian national company, with an average of 197.5 passengers per train over the year. Correspondingly, the highest level of passenger traffic density was also attained by an European railway, 5237 thousand passenger-kms per route-km, by the Dutch operator. In the case of freight traffic, the equivalent figures were 604 tons per train, for the Finnish national operator, and 2819 thousand tons-kms per route-km by the corresponding Belgian company.

With respect to the measurement of productivity, it is often grouped around the productivity indices of labor and capital, as mentioned above. However, since many companies do not detail the volume of employees by activities, often only the aggregate index of total train-kms per employee is usually available. Previous studies (Nash, 1985) have estimated that freight traffic is more labor-intensive than passenger traffic, so this measurement is clearly biased due to the different composition of the output of railway companies. Considering the aggregate index indicative of the volume of train-kms per employee in the case of Europe, the most efficient company in 1994 was the Dutch operator with a volume of 4434 train-kms per worker. In the case of North America, where many companies offer only freight transport, the most efficient companies transported about eleven million tons-km per employee.

Measurements relating to the productivity of capital can be divided into those that refer to traction units or locomotives, or to wagons. With regard to locomotives, an interesting index is the locomotive availability (in %), which indicates the degree of overdue and deferred maintenance and for which Gannon and Shalizi (1995) recommend a value not less than 80%.

Finally, the financial indicators should not be considered of lesser importance by a regulator, even if it is more concerned with operational and commercial performance. For example, the ratio of revenues over total costs may indicate the company's degree of financial solvency, whereas the level of subsidization, and subsidies as a percentage of total revenue or costs, in turn indicates the degree of financial dependence on public bodies. These indicators are very important and should not be independently interpreted since, as shown in Section 2 and by the empirical evidence (see Gathon and Pestieau, 1995, for example) the most heavily subsidized railway companies are often the most inefficient ones.

The main conclusion to be drawn from this section is that performance indicators, despite being very useful, should be designed and interpreted with care. Reference levels and comparisons must only be treated as provisional guides and not given firm normative status. Individual indicators must not be analyzed in isolation from others. There is not a unique optimal level for a single indicator, nor is there an optimal profile for several. An appropriate appraisal requires tradeoffs measuring the relative cost of changing levels of different indicators and the relative importance of the objectives that the indicators reflect.

7. – REGULATORY INSTITUTIONS FOR THE RAIL INDUSTRY.

7.1. – Some issues in the design of a regulatory agency for the rail industry.

The design of regulatory institutions for the rail industry encompasses several issues relating to such a broad number of matters that they are difficult to summarize. According to the incentive literature (Laffont and Tirole, 1993), this subsection proposes a list of four key items that should be considered, at least, at the beginning of such regulatory design: the degree of independence of the regulatory agency; its relationship with the government; the scope and jurisdiction of the regulator; and the number of regulators and its appointment.

7.1.1. – The degree of independence of the regulatory agency.

Given the extensive experience of national railways around the world, the independence debate should be the first and most defining issue to deal with in the design of a regulatory agency in the rail industry. In a broad sense, the term "independence" pertains to the relationship among the regulator and political authorities, regulated firms, consumers, and other private interests. However, this simple definition conveys a common misunderstanding that may shed some confusion on its real significance, since no regulatory entity can be truly independent. Even if a regulatory entity is a non-ministerial commission, it is still a body created by the government. Thus, independence just means that the regulator does not have to get the approval of any political body in its activity (Tenenbaum, 1996). In the case of rail regulation, only a few countries have fully independent agencies (as in the case of the United Kingdom), whereas in most of them, railways are still considered in some sense a strategic sector.

7.1.2. – The relationship of the regulatory agency with the government.

The degree of discretion or autonomy in regulatory processes differs widely among countries and industries, according to their particular features. Most systems lie somewhere between the extreme case of broad discretion (where the laws only define "just and reasonable" behavioral standards) and the case of tightly specified laws or almost-complete contracts between the regulatory body and the government, as in some European countries and New Zealand.

An important aspect to consider, with particular importance to developing countries, is the overall political and social stability of the country, and the historical reputation for respecting private property rights as a function of the development of its legal system. In fact, the higher a country scores on these criteria, the more discretion its regulators can retain without significantly increasing the risk premium of the cost of capital demanded by investors in their rail or infrastructure services.

In any case, to be autonomous, rail regulatory agencies must first have their own resources, since purely relying on government or budgetary transfers controlled by politicians, is a threat not only to their autonomy and effectiveness, but, more dangerously, to their independence. The most common method of funding is through levies on the regulated firms or

the consumers of the regulated services (as in the United States or United Kingdom), where these levies can be viewed as user fees for the coordination services provided by the regulator.

7.1.3 – The scope and jurisdiction of the regulator.

In principle, rail regulatory authorities are usually industry-specific, co-existing with separate agencies for each of the most important utilities, as in the United Kingdom. However, they could also be sector-specific, with separate agencies for groups of related industries, or multisectoral, with a single regulatory agency for all or most infrastructure sectors.

Although the choice of a particular regulatory scope has not followed a single method in most countries, it is agreed that a multisectoral agency offers some advantages over its alternatives, particularly in countries with limited regulatory capacity. In fact, a transport authority (either at the national level or for particular areas or cities) would pool regulatory resources and tend to increase resistance to regulatory capture or political interference. Simultaneously, multisectoral regulation improves the consistency of decisions across sectors and reduces the risk of economic distortions and allows dealing with blurred industry boundaries (like in the case of urban transport).

On the side of the disadvantages, advocates of industry-specific agencies for the rail sector often argue that multi-industry agencies may lack sufficient industry-specific expertise or focus on the particular technology. A second fear relates to the worst consequences associated to the failure (by corruption or incompetence) of a single regulatory entity, and it is finally suggested that multiproduct-industry agencies are only appropriated for small economies with limited regulatory skills.

In any case, the choice of one form of regulation or another must be accompanied by a well-defined list of competencies, since when the frontiers of the regulatory activity of an agency are not fully specified, it is foreseeable that disputes will arise. Whatever the scope adopted for the agency, the rules of the regulation game must be clear from the beginning not only for the regulatory institution, but also for the regulated firms, whose owners and investors demand a safe financial environment to carry out their investments. In Britain, for example, the Office of the Rail Regulator openly publicizes its aims and objectives, so that all the agents in the industry know their jurisdiction and its limitations.

7.1.4. – The number of regulators and its appointment.

Many countries entrust their transport regulation decision-making authority to a commission or a board of several members; others prefer a single individual. Smith (1996) evaluates each of these alternatives according to the desired aims of the regulatory body. Individuals (such as the British Rail Regulator) score better in speed of decisionmaking, accountability and predictability, whereas commissions (such as certain Transport Commissions in South America) are preferable with regard to invulnerability to external influences, potential to reflect broader perspectives, and potential to stagger terms to enhance stability.

With respect to the appointment of particular people to the staff of the regulatory agency, this is a question closely related to their autonomy or discretion. To preserve it, rail regulators should have autonomy in staffing and budgeting, so that they could recruit staff with high levels of expertise. Since overstaffing can lead to interference with the commercial operation of regulated firms, the tasks of the agency should determine the size of the staff, and not political considerations such as how many people have lost their jobs through as a result of privatization. Some appointment criteria, such as particular interests in the industry or inadequate training, are particularly critical for new regulatory agencies that have yet to establish a reputation for competence and reliability.

7.2. – Regulatory institutions in practice: Argentina and the United Kingdom.

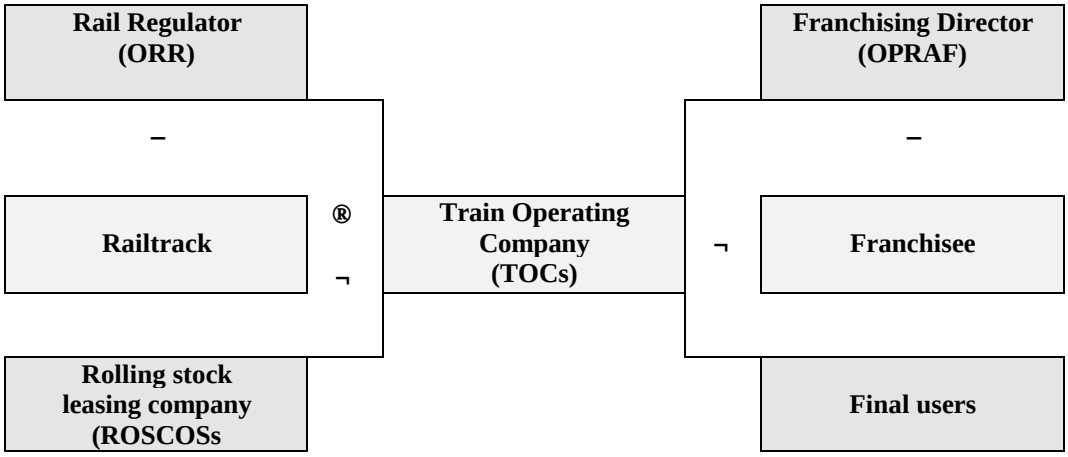
All the features so far described allow us conclude that, with few particular details, the rail industry does not require a particular treatment in the design of regulatory institutions. Some controversial issues, however, can be better understood by comparing case studies to illustrate two different approaches to institutional rail regulation: United Kingdom and Argentina.

The privatization, unbundling and later, the franchising of former *British Rail* in the United Kingdom, constitutes a paradigm widely mentioned in other parts of this paper (Bradsaw, 1996). In fact, the rail industry was the last of the big public utilities to be privatized in Britain. The legal and regulatory framework for this privatization was set up with an Act of Parliament in 1993. Several significant structural reforms were implemented prior to privatization, including the horizontal and vertical break-up of the industry. The antecedents in Europe (in Sweden, where the State retained the control of the network) and in South America (where Chile first introduced a total separation of railtracks and train operations) were surpassed in the United Kingdom, where the process aimed at the complete private provision and management of rail services and infrastructures through corresponding franchises.

Box 7.1 describes the institutional regulation both of services and of infrastructures that followed the franchising. At the top level, two independent regulatory bodies were launched. First, the Office of the Rail Regulator (ORR), in charge of regulating access to the national track via price-cap regulation on access charges and of monitoring the overall functioning of the industry. Second, the Office for Passenger Rail Franchises (OPRAF), with the duty to administer, supervise and regulate the franchising of passenger services.[30] The franchisees, selected through a bidding process, owned Train Operator Companies that were based on former British Rail train operating units and provided passenger services. There are also freight companies, although their importance is lower, and certain unregulated passenger operators. At a similar level to the service providers is Railtrack, the private provider of infrastructure, which charges access fees for tracks and lease stations.

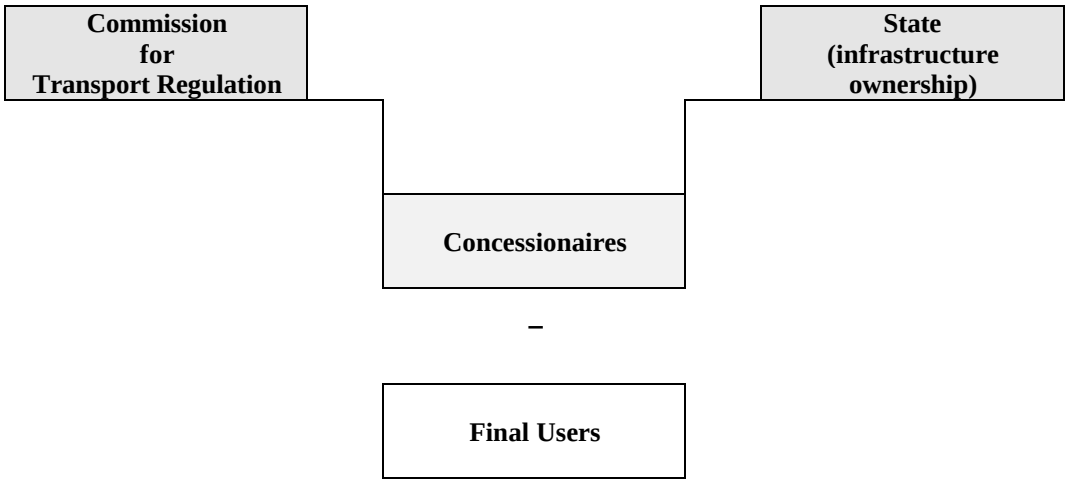
[30] There are also other regulatory bodies not represented in Box 7.1, such as the Secretary of State for Transport (with a supervisory role) and the safety regulator, Health & Safety Executive, as mentioned in Section 4.

Box 7.1. Institutional structure of British rail system.



At the bottom level of the British rail structure are the final users, the passengers and firms demanding transport services, and the auxiliary companies. For example, the existing rolling stock leasing companies provide support to the TOCs, by leasing them locomotives and wagons. They own nearly all passenger rolling stock, and most freight locomotives. Finally, although not represented in the figure, is the sector of infrastructure engineering, infrastructure maintenance units and track renewal units, whose main partner in the system is Railtrack, and which is competitively provided by private companies.

Box 7.2. Institutional structure of Argentina rail system.



The case of Argentina rail restructuring and its later institutional regulation is somehow different, as summarized in Box 7.2. It is characterized by a lower level of complexity and a lesser extent of both private participation and vertical unbundling. It is also interesting because it

demonstrates that infrastructure can be kept in public hands while services, both profitable and social, can be successfully provided by the private sector, thus showing that separation and concessioning can work and yield benefits (Estache, 1997).

The restructuring process initiated with the former public rail operator, *Ferrocarriles Argentinos* (FA), began when it was split into several separate freight and commuter integrated networks that were later concessioned. The infrastructure, however, was not privatized and coordination, particularly in the Buenos Aires Metropolitan Area, was enforced by a unique multi-modal transport authority. Thus, the State remains the owner of the fixed facilities, including tracks and stations, and of the rolling stock. The concessionaires have to pay a fee for the use of the fixed infrastructure and a rent for the use of the rolling stock received from the former FA.

The new regulation of the rail sector in Argentina is headed by a public department, the *Comisión Nacional de Regulación del Transporte* (National Commission for Transport Regulation, or CNRT), dependent on the Ministry of Economy and with jurisdiction over both rail and road freight and passenger transport. This commission is in charge of monitoring the overall performance of the transport sector, including the safety and quality standards. Its competencies also include the design and implementation of the concession contracts.

The comparison between Argentina and United Kingdom rail restructuring processes sheds some lights on the issue of designing regulatory institutions for the rail industry that can serve as a conclusion for this section. The British case is an example of a fully privatized system including both infrastructures and services, with regulatory institutions at different overlapping levels. The Office of the Rail Regulator controls the overall system, but some of the regulatory functions are in hands of the OPRAF (with respect to passenger traffic) and Railtrack (with regard to infrastructure).

In summary, the design of regulatory institutions for any industry, and the rail sector is not an exception, is a complex task that involves the balance of many different forces among which the more important ones are the regulator's objectives and resources, the institutional structure of the country, and the concrete industry conditions faced in each case. Different regulatory objectives have differing direct effects on the type of regulatory controls used in the rail industry. For example, the ultimate reason behind several rail concessioning processes in South America was the proven underperformance of the existing national railways, whereas the privatization of British Rail in United Kingdom also aimed at political objectives.

Similarly, limited resources also affect the nature of regulatory activities. Detailed information about current operating technologies, potential alternative technologies, and consumer preferences is the key to designing effective regulation policy in the rail industry. Substantial information about the performance of the regulated railroad is also required in many settings to implement and enforce regulatory policy, in the sense described in previous section. If regulators have excellent resources (e.g. a large, well-trained and experienced staff to perform research), they are often better able to understand the environment in which they operate and make better policies.

As a second influencing element, the institutional structure of the country in which

regulation is imposed may affect two important aspects of the regulatory process: the regulator's ability to deliver promised rewards or threatened penalties and the set of complementary control instruments used in this environment. Thus, regulatory policy is only effective inasmuch as it influences the activities of the regulated firm. To do so, regulatory policy must create systematic, credible and durable links between the firm's activities and its financial well being. The regulator's commitment ability facilitates the birth of these links and is determined by a variety of factors (such as the political pressure and the independence the regulator may gain from it, particularly from a financial point of view), and the form of appointment of the regulator (number of appointees, disqualifying of candidates, length of term mandates, etc.).

Finally, as the third element of the design of the regulatory activity, it should be considered that the proper scope, form and function of regulation should depend on the concrete industry conditions faced in each case. The type of productive technology, the consumers' willingness expressed through their demand, or the information conditions in the markets is likely to affect the interaction of the regulator and the regulated firms. For example, in the rail industry, when the production technology exhibits increasing returns to scale, regulation should be concerned on how to minimize potential market power threats. If no other suppliers exist, and benchmarking is no longer possible, the regulator may be forced to investigate the firm's operations in lengthy detail before it can fashion reasonable regulatory rulings.

8. – CONCLUSIONS.

The economic theory of transport regulation has undergone an explosion of interest over the last decade. Traditionally, it was viewed as a pure exercise of second best optimization, and largely based upon the acknowledgment of the existence of huge informational problems. Simultaneously, the worldwide rail industry has experienced a deep restructuring process aimed at stopping the erosion of the sector's share in transportation markets. The process adopted diverse forms according to each country's objectives and circumstances, varying from simple reorganization measures to extreme restructurings, in which private participation and vertical separation of infrastructures from service provision, have been the most defining features.

Consequently, this paper suggests that the regulation of the rail industry today cannot remain unaffected by these changes. A number of new regulatory scenarios have emerged and further regulatory issues have appeared, such as the definition of contracts for private participation, the adaptation of traditional price controls to the new environment, the definition of quality surveillance instruments, and the design of mechanisms to manage and plan infrastructure investments in this context.

Some new problems and difficulties have also arisen after the restructuring process. For example, in cases where licenses have been used, several concessionaires have been unable to fulfill the objectives agreed in the concession contract. This is a reason for making the contracts flexible enough to take account of novel situations that may affect the companies' performance. On the other hand, the credibility of the system cannot allow unjustified and systematic deviations from the objectives laid out in the franchise.

Therefore, in all cases, total support of the government for the privatization or deregulating process has been a key factor in order to provide the necessary changes in the law. It

is also necessary for the private sector to become involved in the process, contributing the commercial approach necessary to its success. This is particularly the case in less developed countries and emerging economies, where the privatization of the rail sector seems to have had positive effects on the modernization and development of the industry.

In any event, proper measures from a regulatory point of view must be taken to safeguard interests other than strictly private ones. Thus, the development of the railway industry should constitute a mechanism favoring development and regional equilibrium, as well as maintaining certain public service obligations.

Finally, although the usual belief is that private rail companies are more efficient than public ones, some experiences show that important increases can be achieved in efficiency levels without needing to fully privatize the industry. In addition, there remain some critical views with regard to the final viability of the restructuring experiences in some countries. In particular, it has been mentioned that the new system for the railways is accompanied by a complex institutional requirement that often entails more relationships and higher transaction costs among the agents within the industry.

On the other hand, the vertical separation of infrastructure and services, undoubtedly presents advantages, but there are also disadvantages such as the loss of economies of scope deriving from a network integrated at strategic, tactical and operational levels. It should also be pointed out that the franchising processes do not always ensure a competitive result.

In conclusion, although there is no unique form of rail regulation to face the new challenges, the general behavioral rule and main advice from this paper is to maintain flexibility and simplicity whenever possible. These criteria suggest that the design of license contracts that include private participation, and the organization of the industry adapted to each country's needs and characteristics, should be viewed as two of the key issues in the new regulatory environment of the rail industry. In turn, the use of these mechanisms also changes the role of the rail regulator, whose actions should be, from now on, governed by principles that foster competition and market mechanisms whenever possible, while simultaneously providing a stable legal and institutional framework for economic activity. The regulator should refrain from interventions unless the ultimate goal of achieving economic efficiency subject to the socially demanded level of equity is in jeopardy. This paper was intended to offer some hints on how to carry out this goal.

9. – REFERENCES.

Amstrong, M., S. Cowan, and J. Vickers (1994). *Regulatory Reform – Economic Analysis of British Experience*, MIT Press, Cambridge, Mass.

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.; J. C. Panzar and R. D. Willig (1982). *Contestable Markets and the Theory of Industry Structure*. New York: Harcourt Brace Jokanovich.

Borts, G.H. (1960). "The Estimation of Rail Cost Functions", *Econometrica*, 28, pp. 108-131.

Boyer, K.D. (1997), *Principles of Transportation Economics*. Addison-Wesley. Reading, Mass.

Bradsaw, B. (1996). "The Privatization of Railways in Britain", *Japan Railways and Transport Review*, 30, pp. 15-21.

Braeutigam, R. (1980). "An Analysis of Fully Distributed Cost Pricing in Regulated Industries", *The Bell Journal of Economics*, 18, pp. 182-196.

BIE. (1995a), Bureau of Industry Economics). *Rail Freight 1995. International Benchmarking*, Report 95/22, Canberra, Australian Government Publishing Service.

BIE (1995b), Bureau of Industry Economics. *Issues in Infrastructure Pricing*, Research Report 69, Canberra, Australian Government Publishing Service.

Button, K. J. (1993). *Transport Economics*. Edwar Elgar Publishing Limited. New York.

Button, K.J. (1994), "Privatization and Deregulation: Its Implications for Negative Transport Externalities", *The Annals of Regional Science*, 28, pp. 125-138.

Carbajo, J.C. and A. Estache (1996), "Railway Concessions – Heading down the Right Track in Argentina", *Viewpoint*, 88. Finance and Private Sector Development Department, The World Bank, Washington D.C.

Carbajo, J.C., Estache, A. and D. Kennedy (1997), "Regulating the Quality of Privatized Transport Services", mimeo., The World Bank, Washington D.C.

CEMT (1996). *Evolution des Transports*. OECD Countries (1970-1994). Paris.

De Rus, G (1998). *Price Regulation in Transport*. First International Training Program. Privatization and Regulation of Transport Services. January 5-13. Washington, D.C.

Dodgson, J. (1994), "Access Pricing in the Railway System", *Utilities Policy*, 4, pp. 205-213.

Estache, A. (1997), "Designing Regulatory Institutions for Infrastructure. Lessons from Argentina", *Public Policy for the Private Sector*, 114, The World Bank, Washington D.C.

Friedlander, A., E.R.Berndt, J.S. Chiang, M. Showalter and C.A. Vellturro (1993). "Rail Costs and Capital Adjustments in a Quasi-Regulated Environment", *Journal of Transport Economics and Policy*, 32, pp. 131-152.

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 L.S. Thompson (1993). *Forms of Private Sector Participation in Railways*, Transport, Water and Urban Department Report No. TWU-9, The World Bank. Washington D.C.

Gannon, C. and Z. Shalizi (1995), *The Use of Sectoral and Project Performance Indicators in Bank-Financed Transport Operations*, Transport, Water and Urban Department Paper No. TWU-21, The World Bank, Washington D.C.

Gathon, H. J. and P. Pestieau (1995). "Decomposing Efficiency into its Managerial and its Regulatory Components: The Case of European Railways". *European Journal of Operational Research*, 12, pp. 500-507.

Guasch, J.L. and P. Spiller. (1996). *Managing the Regulatory Process: Design, Concepts, Issues and the Latin America and Caribbean Story*. The World Bank. Washington D.C.

Guislain, P. and M. Kerf (1995), "Concessions – The Way to Privatize Infrastructure Sector Monopolies", *Viewpoint* 59. Finance and Private Sector Development Department, The World Bank, Washington D.C.

Häfner, P. (1996). "The Effects of Railroad Reform in Germany", *Japan Railways and Transport Review*, 15, pp. 27-30.

Institute for Transport Studies (1994), *European Railway Comparisons*, Final Report, ITS Working Paper 418, University of Leeds.

Kerf, M., R.D. Gray, T. Irwin, C. Lévesque and R.R. Taylor (1997). *Concessions for Infrastructure: a Guide to their Design and Award*, Technical Paper, 399, Finance, Private Sector, and Infrastructure Network, The World Bank, Washington, D.C.

Kessides, I.N. and R.D. Willig (1995), *Restructuring Regulation of the Rail Industry for the Public Interest*. Policy Research Working Paper 1506, The World Bank, Washington D.C.

Kopicki, R. and L.S. Thompson (1995). *Best Methods of Railway Restructuring and*

Privatization. CFS Discussion Paper Series, no. 111, The World Bank, Washington D.C.

Laffont, J.J and J. Tirole (1993), *A Theory of Incentives in Procurement and Regulation*, Cambridge, Mass., MIT Press.

Layard, R. and S. Glaister (eds.) (1994), *Introduction to Cost-Benefit Analysis*, Cambridge University Press, Cambridge.

Liston, C. (1997). "Price-Cap versus Rate-of-Return Regulation". *Journal of Regulatory Economics*, 5, pp. 25-48.

Lundberg, A. (1996). "Restructuring of the Swedish State Railways". *Japan Railways and Transport Review*, 218, pp. 22-26.

Mitchell, B. and K.J. Budin (1997), "The Abdijan-Ougadougou Railway Concession", Infrastructure Notes, Transport, RW-14, Transport, Water and Urban Development Department, The World Bank, Washington D.C.

Nash, C. (1985). "European Rail Comparisons. What can we learn?". In Button, K. J. and D. E. Pittfield (eds.). *International Railway Economics*. Gower, Aldershot.

Nash, C.A. (1997), *Rail Privatization. How is it going?* Institute for Transport Studies, ITS Working Paper 497, University of Leeds.

ORR (1997), Office of the Rail Regulator, *Licensing of Rail Operators*, HMSO, London.

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. (1996), "The Economics of British Rail Privatization: an Assessment", *Transport Reviews*, 16, 1, pp. 1-21.

Preston, J., G. Whelan, C.A. Nash and M. Wardman (1996), "The Franchising of Passenger Rail Services in Britain", Institute of Transport Studies, Leeds.

Shaw, L.N., K.M. Gwilliam and L.S. Thompson (1996), "Concessions in Transport", Transport, Water and Urban Department Paper No. TWU-27, The World Bank. Washington, D.C.

Smith, W. (1996), "Utility Regulators: Creating Agencies in Reforming and Developing Countries", (mimeo.). International Forum for Utility Regulation, UK, June 1996. [Reprinted in *Public Policy for the Private Sector* (1997), 127-129, The World Bank, Washington D.C.]

Stigler, J.E. (1987), "The Causes and Consequences of the Dependence of Quality on Price", *Journal of Economic Literature*, 25, pp. 1-48.

Swift, J. (1997a). *Regulating the Railway in the Public interest*. Office of the Rail Regulator. Railway Study Association. Sessional Meeting at the London School of Economics, 11 June.

Swift, J. (1997b). *Regulation of the Railways in Great Britain: 1993-1997 to 2000+*. Office of the Rail Regulator. Address to the European Union Delegation of the French National Assembly and the Senate. Paris, 13 March.

Talley, W. K. (1988). *Transport Carrier Costing*. Gordon and Breach Science Publishers.

Tenenbaum, B. (1996), "Regulation: What the Prime Minister Needs to Know", *The Electricity Journal*, March, pp. 28-36.

The World Bank (1996), " Argentina Transport Privatization and Regulation: The Next Wave of Challenges", Infrastructure Division, Report No. 14469-AR, The World Bank, Washington, D.C.

Thompson, L.S. (1993). "The Evolution of Railroad Regulation in the United States". In Jose C.Carbajo (ed.), *Regulatory Reform in Transport: Some Recient Experiences*. The World Bank. Washington D.C.

Thompson, L.S. and K-J. Budin (1997), "Global Trend to Railway Concessions Delivering Positive Results", *Public Policy for the Private Sector*, 134, The World Bank. Washington D.C.

Thompson. L.S. and J.M. Fraser (1996), "Financial success hinges on productivity", *Database, Rail Business Report 1995*, The World Bank, Washington.

Waters II, W.G. (1985). "Rail Cost Analysis". In Button, K. J. and D. E. Pittfield (eds.). *International Railway Economics*. Gower, Aldershot.