European Transport \ Trasporti Europei n. 29 (2005): 22-45

Recent changes in the global rail industry: evaluating the new regulatory instruments

Pedro Cantos[1*], Javier Campos[2]

[1] *Departamento de Análisis Económico (Universidad de Valencia)*
[2]*Departamento de Análisis Económico Aplicado (Universidad de Las Palmas de Gran Canaria)*

Abstract

The changes faced by the global rail industry in recent years have brought a redefinition of some of the traditional regulatory instruments available in this sector. This paper, focusing on price and quality regulation, discusses how these instruments have been applied in several countries where private sector participation in railways has been introduced mainly through concession contracts, and where some form of vertical and/or horizontal unbundling has been implemented.

Keywords: Railways; Price regulation; Quality regulation.

1. Introduction

After enjoying an unchallenged position for more than 100 years as the dominant means of transport, the rail industry has globally faced a dramatic change both in terms of economic relevance and organizational structure during the last decades. The decline of the railways has been partially explained by the government involvement in its management and the pervasive effects of an obsolete regulatory framework, which impeded, or at least slowed, the necessary adaptation to a changing environment dominated by more flexible transport alternatives.

Narrowly classified as natural monopolies since the XIX century, railways' management around the world widely relied on an undisputed model based on a vertically integrated firm, heavily protected from competition which acted as a national provider of a public service and received generous support from the Government. With very few exceptions, this was the paradigm until the 1980s, when a series of reforms, in the UK, Chile, New Zealand or Japan proved that competition could be introduced in this model through horizontal and/or vertical unbundling, and the subsequent increase in

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

private participation in the provision of services and, although less successfully, in the management of infrastructure.

Within the traditional railway model, pricing and quality decisions were heavily regulated and political interference in managerial decisions often affected these aspects of the railway companies. In fact, pricing rules were relatively simple: in most cases the overall scheme was characterized by maximum prices with little connection with costs, combined with cross subsidization, through which some profitable services pay above their avoidable costs maintaining unprofitable services paying below their avoidable costs. Subsidies, not necessarily associated to public service obligations completed the picture. With respect to quality, few commercial provisions were in practice, since the Government-owned nature of most companies prevented them from making a real effort on improving this issue.

For these reasons, the main aim of this paper is to discuss some of the new regulatory instruments on price and quality regulation that have recently become of common use in the countries which have opted for a change in their railway organizational paradigm. In section 2, we will first review the principles behind the price regulation mechanisms governing the provision of (mostly, passenger) services in a context of a possibly unbundled rail industry company enjoying a significant degree of private participation (usually, through a concession contract). In section 3, we specifically study in detail two of the major problems arising in the regulation of rail infrastructure, provision and access. Finally, since tariff controls can easily be cheated on quality grounds, quality requirements become essential for monitoring overall performance of rail concessionaires. We thus address the issue of quality regulation, including safety concerns in section 4; this includes not only the adequate definition of quality targets, but also a review of the most relevant mechanisms for quality control currently used in the rail industry. The final section describes some performance indicators that could be applied to monitor the behaviour of the regulated rail companies, thus providing a useful device aimed at moving from the definition of the regulation theoretical principles to the problem of how to implement them.

2. Price regulation of rail services: principles and mechanisms

According to standard economic principles, prices for rail transport services should match the opportunity cost of providing it in order to make the most efficient use of the economy's resources. This is the economic efficiency or *first best* criterion which has defined the traditional regulation of the rail industry during the last fifty years. The main focus of government regulation was controlling market power by setting prices that limited the monopolistic abuse of any particular railroad. The exact form of tariff control (official approval of rates with little or no degree of financial autonomy) in each case depended on the nature of the industry, the ownership of the assets, the complexity of the regulated service, and the social and political pressures to maintain financial equilibrium in the medium and long run.

In practice, however, opportunity cost pricing implies some measurement difficulties and often conveys economic losses, especially in industries with large economies of scale (Armstrong, Cowan and Vickers, 1994). Therefore, this form of regulation was complemented by a number of standard price mechanisms that economic theorists devised to substitute the ideal efficiency criterion of pricing each unit of service at the exact cost of its provision.

Price discrimination policies, either by type (student and senior prices, frequent traveller and commuter passes), number of consumers (group discounts), type or volume of freight (cargo rebates for some goods) or time of day or season (peak-load prices), have always been common in transport. The use of two-part tariffs, with fixed and variable components, is also a common tariff policy in which each unit of consumption (for example, a single trip) is priced differently. These mechanisms allow greater flexibility for railways and increase revenues without a great effect on costs. However, their social acceptability and information requirements can limit the extent of their application.

In the new regulatory environment defined by the changes experienced in the rail industry since the 1980s, where separation of the infrastructure from services has been widely implemented in diverse forms, and a notable degree of private participation in rail management exists through, for example, concession contracts, pricing principles must be put into practice by means of concrete rules within the contract itself.

In general, as private operators, rail concessionaires are allowed to set prices relatively freely, price regulation has a different nature: instead of price-setting, it becomes more price-supervision. To carry out this task, most concession contracts awarded in the rail industry (for example, in Argentina or Brazil) routinely include a specific procedure to control and evaluate the prices set by operators. These price control mechanisms are generally set according to three key factors: *(i)* the degree of monopoly power effectively conferred to the operator; *(ii)* the extent of government non-commercial objectives in the concession award procedure; and (*iii*) the possible existence of other limiting factors, such as intermodal competition. This latter element is relevant in rail freight operations (intermodal competition from trucking), but in the case of passenger traffic (especially commuter and regional), social pressure for low fares usually dominates many price interventions. In practice, the most common alternatives (*second best criteria*) for price control in rail concessions adopt the form of a rate of return regulation or a price cap mechanism.

2.1. Rate of return mechanisms

Rate of return regulation is 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. The regulator typically determines a revenue requirement based on a firm's 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 base rate (Liston, 1997).

Thus, rate of return regulation has three components: the base rate, the allowed rate level, and the rate structure. The base rate 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 base rate may either jeopardize the survival of the firm or allow it to earn excessive profits.

In practice, the base rate usually includes most fixed costs less depreciation and working capital.

Three characteristics should govern the definition of the asset base rate. First, with respect to the treatment of past investments carried out by the railroad before the regulatory period,[1] it should be consistent and transparent in order to ensure that assets are not expropriated *ex post* by opportunistic regulatory behaviour, which would increase the cost of capital required by investors. Second, with future investments and expected operating expenditures and costs should be considered in the asset base definition inasmuch as they do not imply "excessive" investment and only when they are fully incorporated into the firm. Finally, with respect to current investments, a 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 that existing assets are sunk, so the opportunity cost of using them in their present activity is zero. If the regulator seeks maximum efficiency, it should ensure that the rate of return structure (and, indirectly, the prices) are set to cover future avoidable costs.

Since most of the assets currently used by railways are financed before the concessioning process, both of these solutions are troublesome. Market values are much lower than replacement costs so this valuation would yield large price increases and windfall gains for private shareholders at the expense of consumers. On the other hand, in attributing a zero value to the existing assets, windfall gains would go in the opposite direction and the proprietors would be reluctant to finance future investments with such a lower real return. A possible way to address this problem is to use some average procedure that considers either a financial projection of what will happen with the future base rate or calculates indicative values by estimating the cash flows that the firm would have earned had the regulatory regime remain unchanged.

Despite its advantages within the traditional price regulation mechanisms (mainly its 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 on to 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;[2] and, finally, the high degree of discretion enjoyed by the regulator in determining the base rate and the rate of return reduces the incentive for rent-seeking behaviour on the part of the regulated firm.

2.2. 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 cost reimbursement rules that firms themselves select according to their preferences for sharing operating costs with the regulator. The basic aim of these mechanisms is the achievement of dynamic efficiency (in the sense of the regulated firm achieving the lowest unit cost in

[1] This is often the case in many restructuring processes when a former state-owned railway transfers its assets to private concessionaires.

[2] 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.

the long run) by sharing some of the efficiency improvement rents between the firm and the regulator.[3]

Alternatively, price cap regulation is another incentive used in both railways and other privatized utilities. In its most standard form, it consists of setting traditional maximum price schemes based on 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 firm's future increases in efficiency. This mechanism came 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 information rents that it implies.

There are a number of minor variations of the price cap system. In the rail industry, one of the most developed is the RPI-X formula. In this setup, the price for a basket of the firm's prices can increase in any one year by no more than the increase in the retail price index (RPI) for that year, minus some fixed-cost (efficiency related) parameter X. In the case of multi-product activities, such as railways output, 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 (calculated with accounting figures) can grow at most by RPI-X. Thus, the regulator can control the prices of multi-product firms by focusing on their revenues and correcting them according to adequate weights. It starts with a reference price, often calculated with rate of return criteria, and set the price for a fixed number of years.

In the United Kingdom, for example, the price cap mechanism, in its RPI-X formula, has been applied to passenger traffic franchises. Commuter fares are regulated with respect to a basket containing all relevant fares, weighed broadly by the income that the operator derives from each. For 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; after January 1999, the price cap was set at RPI-1%.

The goal of this method is to increase the efficiency of the regulated rail operator, allowing the firm to earn substantial profits by improving 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, future investment programs, 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 mechanism. The cost of capital and the value of existing assets are 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

[3] There are several ways to accomplish this goal and implement its results. For example, the sliding scale plans used in the United Kingdom's Railtrack 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.

effect of the price cap on the future shape of the market is 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 of using indexes or indicators (as described below) to measure the difference between aggregate rates of growth of outputs and inputs and therefore calculate productivity from the residual. Econometrics also provides alternatives for estimating cost functions and their corresponding productivity parameters. Once the X-factor is determined, the initial price ceiling that is imposed on the firm after a switch of regime is critical. If the caps are too high, then too little surplus is transferred to consumers and deadweight losses are huge. If they are set too low, the firm may not be able to break even and may then have difficulty attracting capital, leading to a deterioration of quality of service.

Another important element of RPI-X regulation is the existence of cost pass-through provisions, through which the firm can transfer to customers unexpected increases in certain factors outside of its control. Although these clauses are standard in the regulation of other utilities, they are not 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 pass-through onto customers could be established in the concession contract.

3. Regulation and rail infrastructure

After reviewing the principles and mechanisms of price regulation for rail services, this section addresses the two most relevant problems of infrastructure regulation nowadays in a context of vertical unbundling and private participation. We first focus on the recovery cost problem and then study the issue of access pricing.

3.1. How to recover infrastructure costs?

Rail infrastructure provision and management are characterized by a high ratio of fixed to marginal costs, the existence of avoidable costs and unavoidable or common costs. Avoidable costs are uniquely associated with a particular output: if this output is not produced, no cost is incurred. 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 result in large deficits that jeopardize the long run survival of the firm. 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 (Talley, 1988).

The existence of cross-subsidization problems in pricing rail services or infrastructure produced in the presence of common costs can be illustrated with the case of a profit-regulated railroad connecting two large cities and also providing rail service to a smaller town along the route between the two cities. The fares charged for passage from the small town generate revenues exceeding the additional cost 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, signalling, and trainyard costs. The issue is how to allocate common costs among customers and services. In many cases, cost sub-additivity and efficiency require 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 relevant issue for efficient pricing of infrastructure 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, 1989). The arbitrary nature of fully distributed cost methods and its lack of a conceptual foundation 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 tests do not apply if railroads are not allowed to abandon unremunerative facilities or services (Kessides and Willig, 1995). If that freedom is denied, a railroad cannot earn adequate revenues if its rates on potentially remunerative activities are constrained by stand-alone cost ceilings.

The cost recovery principle should be a central issue in the design of any rail infrastructure pricing procedure. The theoretical and political debate focuses on two options. Many public firms still advocate the use of the efficient price mechanisms 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 thought that it might yield inefficient outcomes for the theoretical efficiency principles, it constitutes the second best available alternative in most cases.

Similarly, with respect to access pricing of 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 sustained in the long-run.

Several econometric studies have shown that in the case of the rail industry, the marginal cost of those railways that are still vertically integrated 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 (see, for example, Friedlander *et al.*, 1993). Price discrimination, if feasible and politically acceptable, may help to raise 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.[4] 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 is to advocate a balance between the cost recovery issue and the efficient pricing rules, giving preferential treatment to one or the other according to the case. However, 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 of investment loans. This continues to be the case in many of the countries of Central and Eastern Europe.

3.2. The access pricing question

The development of tariffs for accessing rail infrastructure varies greatly among different 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 favour the entrants.

Depending on the discretion allowed to the integrated firm, potentially distortive effects on access prices can be determined in several ways. First, when infrastructure is still publicly owned or managed, the regulator can determine the price as an integral 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 regulatory schemes, usually rooted in incentive-based price regulation mechanisms (to favour the firm that achieves higher levels of efficiency). Thirdly, the firm may have discretion over aspects of access pricing subject to some overall regulatory constraint,

---

[4] Ramsey pricing refers to charging higher 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.

and finally, the firm may have full discretion over the price and is only restricted by the country's anti-trust law.

In all of these cases, there are two main approaches to setting access prices when the principles of cost recovery plus the normal rate of return are required. First, some countries use the current dominant paradigm for setting access charges: *cost-related charges,* which are based on the optimal first-best principle of pricing according to marginal cost (considered the forward-looking long-run incremental cost). The higher the proportion of common costs, the more complex the principle. It is based on the so-called *efficient-component rule*, which determines that optimal access charge is equal to the direct cost plus the opportunity cost of providing access (given by the reduction in the dominant firm's profit). To compute these costs, the regulator has to consider economic depreciation (physical depreciation plus technological progress) and forecast future usage.

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 the privatization and regulation methods and the extent of competition envisaged in bidding for the right to operate concessioned infrastructure services, the former is more likely to reflect a past situation that domestic reforms are trying to overcome.

The second method of setting access prices consists of developing *usage-related* charges. Once-avoidable costs are covered by increasing prices that are inversely related to demand elasticity. Another option (less controversial) is the use of a two-part tariff to avoid service cuts by train operators to save charges even when the network has no cost saving. The British infrastructure provider until 2001, Railtrack, is a well-studied example of how access prices functioned in practice. In a industry context where operating companies were franchised, Railtrack managed the infrastructure (track, signalling systems, electric power supply and stations) and was responsible for its maintenance, new investments and train operations (timetables, coordination, etc.). It also sold access to infrastructure to passenger and freight operators.

Railtrack owned the rail network and set track charges that had to be agreed upon with the rail regulator under the criteria openly published in a number of regulatory policy statements. The price control system operated through a simple RPI-X formula that was revised every five years, remaining fixed between revisions. The structure of Railtrack's access charges for passenger services was based on the usage-related charges and was made up of multiple-part tariffs with at least four elements. 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, to 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 indicated the long run costs imposed on Railtrack in delivering the total access rights of a train operator. Finally, common costs, as the remainder of the fixed charge, designed to recover the rest of Railtrack's costs at the sub-zonal, zonal or national level. This was 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. The first two elements amounted on average to only about 9% of total track access charges, and given the current structure of charges, these were the only elements that vary directly. The remaining 91% of the aggregate charge was in the form of a fixed charge, which did not vary with the number or type of trains run or with passenger revenue.

In the case of freight services, access prices were more flexible. The rail regulator had simply established several principles to be considered by Railtrack in its relationship with private operators. First, prices must cover the avoidable costs incurred by Railtrack as a direct result of carrying that particular freight flow; second, prices must be lower than the stand-alone cost that would be incurred by a national efficient competitor; third, no undue discriminatory charges are possible; and finally, charge structure 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 of infrastructure management and transport operations. Directive 95/18 regulated the licensing of railway undertakings, and Directive 95/19 established several general principles on the allocation of railway infrastructure capacity and infrastructure fee charges. These principles were designed to ensure an optimum, non-discriminatory use of infrastructure and guarantee an access charging policy according to EC rules, but they were received by member states with various degrees of enthusiasm. The objective of most governments that have set rules for infrastructure fees is to cover costs and differentiate fees to reflect different cost factors. In 2001 a new rail package clarified the principles on which rail infrastructure management should be based on and, very recently, through EC Directive 2004/51, the deadlines for implementing ‘third party access’ have been shortened (to January 2007). However, many of these changes have been slowly implemented in most countries.

In France, for example, several principles were introduced to give access to railway infrastructure to licensed international groupings of transport services and operators of combined transport, but present arrangements seem more inclined to promote conventional international rail groupings rather than new entrants into the rail market. With centrally planned timetables, only the domestic operator pays a fixed amount to the (also public) infrastructure manager. User fees are fixed, accounting for a wide set of criteria including: infrastructure costs, the transport market situation, supply and demand characteristics, imperatives based on optimized use, and standard conditions for intermodal competition. In 2004 the access fee system was changed in several important ways. First, the access fees per unit of traffic were set two years in advance instead of essentially being negotiated after the fact. Second, the structure of the fees was changed to sharply increase fees for local passenger trains, freight, and ancillary services (such as stops in stations or the use of marshalling yards). Third, the projected total volume of fees was forecast to increase more gradually at a rate of about 300 million euros per year.

Similarly, in Germany, the federal government owns the track infrastructure and is responsible for its preservation and for securing a certain level of public transport service by means of the *Deutsche Bahn* (DB), an independent joint-stock holding whose sole shareholder is the state. The infrastructure division of DB bears operating and maintenance costs and is in charge of stations, ticket sales, passenger attention, etc. It is also responsible for setting charges for track usage, which are supposed to cover all infrastructure costs, including investment. These charges are based on prices per train/km on the different line sectors, resulting in a number of different fee combinations (Häfner, 1996).

In Spain, the 2003 Railroad Law introduced charges for the usage of rail infrastructure, stations and other track elements that conform to EC Directive 2001/14. These charges intend to recover infrastructure's full costs, and include four components: access, capacity reservation, circulation and traffic. The access charge is a general payment to be made by all licensed operators for the right to use the infrastructure. The capacity reservation and circulation fees depend on the kilometres of track used and vary with the type of service or train, the hour of the day and the characteristics of the track. Finally, the traffic charge is levied on the operators depending on the economic value of their service as measured by the number of seat-kilometres or ton-kilometres operated.

4. Quality and safety 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 achieved by train operators and particular features in regard to 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 the standard price regulation mechanisms described in the previous section. Economic theory provides a well-known argument to answer these questions: real world transport activities are characterized by market failures due to information problems.

In an ideal world with a large number of competitive rail transport service providers and well-informed consumers of passenger and freight services, quality regulation would not be required since market forces would adjust consumer demand (in terms of prices, levels of output and of quality of service) to firm supply. If no price correction took place, less reliable rail companies would be driven out the market and only those whose price-quality ratios were in accordance with demand would remain. However, when full information doesn't exist, markets cannot exert this disciplinary role on firms and purely competitive solutions do not always positively affect quality, prices, or output. Pure competition may result in unsafe, unreliable or unpleasant services since limited availability of resources and lack of adequate control mechanisms make it impossible to adjust consumer and producer interests.

In the traditional organization of the rail industry some years ago, a monopolistic structure with a single firm providing services at the national or local level, price-quality adjustment problems may have increased since the monopoly's privately optimal level of quality may not have coincided with social standards. Simple price regulation is seldom a solution. Any regulated, multi-product monopolist in an environment of asymmetric information tends to degrade quality in order to achieve higher profits once it enters the market. Railway firms are not immune to this temptation, for example, in terms of punctuality and cancellation standards. The quality outcome of any monopolist, not just in the rail sector, heavily depends on the specific regulation adopted. For example, with rate of return regulation, over-investing in non-required technological quality may accentuate the Averch-Johnson effect. Alternatively, with price cap regulation, a subtle cut in quality can be a very tempting way to cut costs (Carbajo, Estache and Kennedy, 1997).

Therefore, the price regulation mechanisms analyzed above are considered incomplete if they do not include quality provisions. This is not always easy, since adjusting price mechanisms by quality may render them inoperative or excessively difficult for the firm to manage or the regulator to monitor. Therefore, most regulators set quality standards or targets for train operators instead of correcting price control mechanisms.

4.1. Definition of quality targets

In setting up those quality standards incorporated in concession contract designs, the regulator often uses the principles of yardstick competition.[5] These quality standards may be constructed at the national or regional level with inter-industry comparisons (as in Brazil and Chile for many of their public utilities) or by establishing international benchmarks or best practices (as in Australia for transport services and infrastructures).

Three elements are considered in detail when designing this process. First, as in other transport modes, quality is mainly measured in service levels or specified service standards. However, this measurement is suited more for factors such as train punctuality, the reliability of services and the waiting time at stations or platforms, than it is for other factors.[6] Simultaneously, the services provided before the transport itself, such as ticketing, reservations, and luggage or cargo handling are often ignored as part of the rail industry's value chain, 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 transport service that includes not only the ride itself, but all aspects related to infrastructure (track and stations), stations and pre- and post-transport services provided to clients.

A second aspect of quality regulation that is particularly relevant to railways is the flexibility with which scheduled services can be changed and new services introduced in response to changes in demand. Here, the rail industry has always been at a disadvantage to road transport because of the need to coordinate working timetables and operations with certain technical requirements due to the lack of alternative routes between points.

Hence, it is not usually easy (with a few increasing exceptions in many countries) for rail transport to offer on-demand services to passengers (for example, as done by charter airlines) or freight customers (door-to-door services). Thus, coordination is relevant for quality of service regulation within the rail firms, and must also be considered in the design of the industry structure. For example, one potential disadvantage of the split between infrastructure and operations is that coordination might be even more difficult when changes have to be negotiated between different organizations, especially where timetable approvals also need to be secured from other train operators using conflicting train paths.

Intermodal coordination with other industries is also necessary, since social quality performance is always evaluated in relation to feasible alternatives. Saturated corridors

---

[5] This is done to avoid the problem of regulator's capture and the discretionary nature of the regulatory action. However, there is a risk of making undue comparisons between different rail systems.

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

(where investment in roads, railways and airports clearly overcomes demand) are a waste of resources that few economies can assume. This almost general equilibrium approach constitutes the third element of the quality regulation process, although it 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 the overall targets to be established in each industry.

Table 1: Quality dimensions of the rail industry.

Dimension		*Definition*	*Measurement Variables*
Quality of Service	Vehicle	Onboard 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

Taking into account these three characteristics, Table 1 summarizes the five most important quality dimensions for the railway industry (vehicle, route, service, social and dynamic quality) along with a number of standard performance measurement instruments for them. The first three (vehicle, route and service) are related to what is usually named (internal) *quality of service*, whereas the last one refers to externalities.

4.1.1. Quality of service

Regulation of the quality of rail transport services in regarding vehicle quality, the transport service itself (aboard trains) and the pre- and post-transport services has been dealt with in different depths in different countries although 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 on the industry post-restructuring.

In general, countries where the sector is still heavily dependent on government or public agencies (such as in Eastern Europe and Asia) have done less to establish separate quality control frameworks than in those where private participation has been significant (for example, the United Kingdom) and detailed quality control systems have been set up. 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 the preservation of incentives to maximize customer service (see Swift, 1997a, 1997b).

This is particularly relevant to the separation of infrastructure and operations. Vertical unbundling in railways distances infrastructure management from the end-user customer and could yield undesirable side effects or contradictions. For example, the density of traffic (trains per day) that maximizes returns on infrastructure investment is likely to be greater than the optimal level from the operators' point of view. This is because at high densities, passenger service is likely to suffer due to congestion. Therefore, no matter whether the separation is institutional or only financial, mechanisms to compensate infrastructure units that run below optimal capacity must be incorporated into contracts in order to maximize end-user customer performance as a whole. Since the particular characteristics of the rail industry in each country require fine tuning of any regulatory or contract enforcement mechanism, Table 2 proposes a simple scheme that identifies and separates the roles to be assigned to the regulator and the operator (either franchisees or public or private monopolies) with regard to quality of service regulation.

Table 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, etc.	✔	×	×
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	✔	×	✔

After its reform and the full privatization of services and track provision, the United Kingdom's rail system constitutes one of the most practical examples of a detailed quality of service regulatory framework (see Table 2). For example, in the case of passenger transport, the regulatory agency (Office for Passenger Rail Franchising, OPRAF) defined what level of service is tendered for particular routes and corridors and 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 franchises operated a poorer service than specified then the OPRAF had the right to withhold the license.

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 the basic provision of services. However, in order to avoid excessively limiting the freedom of the operators, these requirements do not specify detailed timetables for each route, but instead set parameters within which each company must design its own timetable. Passenger service requirements are set out by route and are largely based on the former British Rail 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 number of train cancellations and, where applicable, the level of capacity that needs to be provided. These limits apply in any 28-day reporting period, with three levels determined: *(i)* a call-in level, where OPRAF reviews the performance of the operator; *(ii)* a second level, where the operator is in breach of the franchise agreement, and *(iii)* a third level, which can trigger default of the agreement.

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

In practice, not all of the quality dimensions defined in Table 1 can be incorporated in the same proportion to any service quality mechanism. The British system mainly focuses on the route dimension and is based on their extensive experience with deregulation. When the role assignment proposed in Table 2 is not considered, or its components cannot be easily separated, several quality regulation failures may arise. The most important is the failure to define adequate 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 incentive to provide the optimal price-quality ratio in favour 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, including all issues related to safety and externalities (pollution, congestion, etc.) must also be considered, and it specifically differs from level of service quality regulation in at least four aspects.

The first element is the scope of regulation. Since non-compliance with social quality standards may affect users and non-users of transport services, these standards should always be exogenously set, by national or supranational legislation with intermodal implications, in the case of the rail industry. This is not always the case for timetables, load factors or vehicle size, variables that 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 was envisaged 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 is that a regulatory approach must be used in the latter. Since the risks associated with accidents or potential environmental damages 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 coordinate safety and reliability. This coordination is particularly important when firms move from a public to a deregulated system. Furthermore, in the rail industry, separation of infrastructure from services and the introduction of open access have made it necessary for a rail track controller to ensure safe coordination between different operators who are using the same tracks or stations.

Again using the British railway system as an example, their safety regulator is the Health Safety Executive (HSE), which informs and advises 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 and 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 operational and technical standards to ensure safety and safe interworking in Railtrack's infrastructure.

The third aspect of particular interest to safety regulation in the railway industry is the assessment and assignment of risk. Given the inherent difficulties associated with strict monitoring, incentives exist for quality-regulated private providers of rail transport services to place compliance with safety requirements below the attainment of financial objectives.

In fact, despite recent tragedies, railways traditionally have a good reputation for safety, a perception that converges with statistical proof in most countries. Therefore, one could conclude that safety levels and management are quite sufficient and no particular safety precautions or measures should be taken. However, public outcry, negative social effects and adverse public opinion from a single catastrophe, together with the persistence of regular fatalities (staff accidents, passengers joining and alighting trains, etc.) make it impossible for the regulator to avoid designing measures and policies to diminish individual and social risk.

One of these policies relates to the compulsory insurance against third-party liability, since it may correct the operators' incentives to take excessive risk. In Europe, for example, Directive 95/18 required that operators of train services must 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: the 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 liability for licensed activities. The type, cover, level and identity of the insurer need the approval of the regulator, who sets guidelines on minimum insurance requirements that operators must meet. The operation of licensed activities without insurance approved by the regulator is considered a breach of the license.

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

For example, air pollution is one of the most regulated areas in the road and air transport modes, but is not a critical issue in the rail industry though, there are some notable exceptions in certain countries and routes. Noise pollution in suburban neighbourhoods, areas close to stations and depots and delicate countryside ecosystems has attracted more attention from both the public and regulators. Most countries, therefore, incorporate into 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.

The final issues related to environmental regulation are measuring, analyzing and predicting the emissions of chemical substances (heavy metals, lubricants, dust, etc.) where railway lines are present and assessing the risk to the safety of local residents as a result of rail-related activities (transport of dangerous goods, explosions, etc.) In these cases, 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 national and supranational environmental standards. In Europe, for example, there are EC Directives on air pollution from vehicles that specify environmental standards for vehicle engines and fuel qualities which apply to both vehicles (wagons, locomotives) and transport operations.

4.2. Instruments for quality control

Once the objectives for service and social quality are well established, the next step in devising a quality regulation system for railways is designing control instruments. In principle, there are three alternative mechanisms for regulating quality in the rail industry.

First, the firm can simply be required to publish and report measures of quality every pre-defined period. 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 regulatory process, access to public information is a very delicate issue since it can serve as a disciplinary device for the rail provider and as a strategic instrument to undermine or strengthen the ability of the firm to survive in the market.

A second quality control mechanism is including a direct, 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 obtain authorization to carry out certain technological improvements in order to avoid overinvestment and 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 defined to avoid changes in quality (and consequently, cost reductions) that could be used by the regulated firm to increase profit, even if the same price caps are maintained.

The third mechanism that can be used to control quality is a customer compensation scheme, where grants or payments are awarded to people affected by non-compliance with quality standards. In practice, these mechanisms only work if quality failures can be easily verified. This requires a detailed regulation not only of quality standards, but also of monitoring rules and guarantees for both the regulator and the regulated that the inspection process will be transparent and objective. Moreover, if the compensation is distributed to consumers, either directly by the firm 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 instead specify minimum quality standards for certain parameters of the rail industry, backed by explicit legal sanctions that may include fines or the revocation or withdrawal of the operating license.

Table 3: Instruments for quality control in the privatized 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 of price-control Mechanisms	- Rate of return regulation vs. Price cap Regulation
	- Information revelation obligation	- 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, 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 is often the most difficult task in the design of the quality regulation process. The approach followed by most countries is outlined in Table 3, with a summary of the most important instruments. Thus, the quality regulation process consists of three stages. First, before entering the market (Stage I), the goal is to anticipate and minimize future conflicts between the regulator and the concessionaire.[7] 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 and suburban trains, vehicle

[7] 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 the previous section.

capacities and punctuality can also be set. Finally, in order to not 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 quality incentives in price-mechanisms, to establish the firm's obligation to reveal information and the auditing (external or internal) processes to be carried out. In most cases, the use of technical control instruments (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 according to any of the schemes described above. 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 behaviour of the operator in the market.

5. Performance indicators

Performance indicators are used in the rail industry to monitor the behaviour of one or more regulated firms in order to evaluate the effectiveness of the regulatory measures to which they are subjected.[8] The main advantage of these indicators or indices is that they provide a periodical assessment and control of the firm's activity and continuously update information, simply, quickly, and at a relatively low administrative cost for the regulator.

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

Comparisons across companies usually provide interesting, persuasive results that can help the regulator set objectives and design 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 differ among companies. For example, countries with very liberalized frameworks in their rail industry (the United States, for example) could set desirable productivity indicator levels (or quality of service) that clearly differ from the levels in other more regulated frameworks (such as in Europe).

Similarly, simple indicators should be carefully interpreted over time to avoid contradictions and inappropriate measurements. For example, when assessing railway output, the number of trains/km may be relatively high, while passengers/km or tons/km may be relatively low (if the firm specializes in one type of traffic). Given this conflict, overall performance can be ambiguous. The most practical solution is to jointly interpret

---

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

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 both financial objectives, reflected in a high cost recovery rate, and objectives based on the maintenance of low prices.

Thompson and Fraser (1996) point out that monetary and productivity variables should be carefully defined for inter-firm comparisons. Fares, wages, outputs and inputs vary widely among countries for a large number of reasons that are not necessarily 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 (each with a different price). 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 freight tariff mistakes include not accounting for the different mix of commodities, size of shipment or length of haul. The latter also affects passenger traffic and is particularly relevant since some costs (ticketing, billing and station maintenance, for example) are fixed with respect to the length of the trip but vary with size or distance.

These difficulties are increased when 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, the number of passengers/km or tons/km per employee,[9] depends on such diverse factors (e.g., regulatory environment, structure of the labour 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 account for the different relative environments of each company. Oum and Yu (1994), for example, estimated different efficiency levels for a sample of OECD railway companies by introducing internal factors (such as the characteristics of outputs) and external factors (difference in the legal and regulatory framework between companies).

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

Several external factors that vary widely from country to country and firm to firm substantially influence comparisons. Therefore *contextual indicators* assist in comparative analysis and define desirable performance levels. They include social and economic characteristics of the railways as well as other elements associated with the economy as a whole. Directed mainly at the regulator, they control for the exogenous factors in inter-firm and intra-firm comparisons. Table 4 presents several examples from international statistical sources.[10] Simultaneously, there are many indicators (particularly those for prices and quality of service) that are informative to transport users and provide input for the regulator's control tasks. Jointly with the contextual indicators, these *management indicators* provide the necessary instruments to judge the management and behaviour of the company, and can be grouped at three different levels, summarized in Table 5.

[9] 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.

[10] 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.

Table 4: 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 trains/km Freight trains/km Passengers/km Tons/km • Revenues Passenger revenue Freight revenue • Network indicators Length of line Length of track Electrified track (%) Route/km/km^2 • Density and service Train routes/km per capita Trains/km per routes/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)

Some final practical rules that could be helpful in this process are as follows: *(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)* multiple objectives can be addressed by multiple indicators (jointly interpreted); and finally, in order to assure the utility of the indicators, *(iv)* appropriate data must be provided and *(v)* the management of the indicators' information should be part of the regulatory process.

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

Table 5: 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 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	• Labour productivity Passengers/km per employee Tons/km per employee Passenger trains/km per employee Freight trains/km per employee Total trains/km per employee • Capital productivity Number and kms. travelled 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

Indicators of service quality that were earlier presented should serve the same way 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 implementation of control systems for prices and service quality may limit the firm management and reduce operability. Placing an emphasis on price control or service quality depends on whether it prefers to offer services at the lowest possible price, or offer services with certain standards of quality. All of these indicators allow the regulator to monitor the operators' activities as defined in Phase II of Table 3. Unjustified or systematic breaches of quality standards (insufficient number of trains per hour, lack of punctuality, unreliability, very high indices of load factor, etc) should be accompanied by an appropriate system of penalties, as described above.

6. Conclusions

This paper has reviewed the theoretical foundations to regulate prices and quality service levels in the rail sector. Also the paper describes how the current changes in this industry have provoked the necessity to modify the old mechanisms to control prices and quality decisions. The institutional separation between infrastructure and operations, the horizontal unbundling process and the increasing contribution of private participation have promoted the introduction of novel and new regulatory systems. But the definition and the application of these systems present problems and difficulties which must be appropriately evaluated. The elaboration of a suitable list of measures or indexes which allow to monitor the performance of the industry is crucial in order to reduce these problems.

In conclusion, there is no unique form of rail regulation to address these new challenges, but the general rule is to maintain flexibility and simplicity whenever possible. Two key issues in the new regulatory environment of the rail industry are that private participation is included in license contracts and the organization of the industry is adapted to each country's needs and characteristics. In turn, using these mechanisms also changes the role of the rail regulator, whose actions should now be governed by principles that foster competition and market mechanisms and simultaneously provide a stable legal and institutional framework for economic activity. The regulator should refrain from intervention unless the ultimate goal of achieving economic efficiency subject to the socially demanded level of equity is in jeopardy.

References

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

Braeutigam, R. (1989) "Optimal Policies for Natural Monopolies", *Handbook of Industrial Organisation*, Vol. II.: North Holland.

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

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

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

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

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

Oum, T. H. and Yu, C. (1994) "A Comparative Study of OECD Countries' Railways", *Journal of Transport Economics and Policy,* 38, pp. 121-138.

Swift, J. (1997a) *Regulating the Railway in the Public Interest.* Office of the Rail Regulator. Railway Study Association, 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, 13 March: Paris.

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

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

Acknowledgements

Pedro Cantos thanks financial support from Spanish Ministry of Science and Technology under project SEJ 2004-00110. Javier Campos gratefully acknowledges financial support from the Spanish Ministry of Science and Education and from FEDER through grant SEJ2004-00143/ECON. Both authors are grateful to The World Bank for collaborating in previous versions of this paper.