OECD
ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

ECMT
EUROPEAN CONFERENCE OF MINISTERS OF TRANSPORT

TRANSPORT RESEARCH CENTRE

TRANSPORT SERVICES: THE LIMITS OF (DE)REGULATION

ROUND TABLE

129

TRANSPORT RESEARCH CENTRE

REPORT OF THE
ONE HUNDRED AND TWENTY-NINTH ROUND TABLE
ON TRANSPORT ECONOMICS

held in Paris on 13th-14th May 2004
on the following topic:

TRANSPORT SERVICES: THE LIMITS OF (DE)REGULATION

OECD
ORGANISATION FOR ECONOMIC
CO-OPERATION AND DEVELOPMENT

ECMT
EUROPEAN CONFERENCE
OF MINISTERS OF TRANSPORT

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

The OECD is a unique forum where the governments of 30 democracies work together to address the economic, social and environmental challenges of globalisation. The OECD is also at the forefront of efforts to understand and to help governments respond to new developments and concerns, such as corporate governance, the information economy and the challenges of an ageing population. The Organisation provides a setting where governments can compare policy experiences, seek answers to common problems, identify good practice and work to co-ordinate domestic and international policies.

The OECD member countries are: Australia, Austria, Belgium, Canada, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The Commission of the European Communities takes part in the work of the OECD.

OECD Publishing disseminates widely the results of the Organisation's statistics gathering and research on economic, social and environmental issues, as well as the conventions, guidelines and standards agreed by its members.

EUROPEAN CONFERENCE OF MINISTERS OF TRANSPORT (ECMT)

The European Conference of Ministers of Transport (ECMT) is an inter-governmental organisation established by a Protocol signed in Brussels on 17 October 1953. It comprises the Ministers of Transport of 43 full Member countries: Albania, Armenia, Austria, Azerbaijan, Belarus, Belgium, Bosnia-Herzegovina, Bulgaria, Croatia, the Czech Republic, Denmark, Estonia, Finland, France, FRY Macedonia, Georgia, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Moldova, Netherlands, Norway, Poland, Portugal, Romania, Russia, Serbia and Montenegro, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, Ukraine and the United Kingdom. There are seven Associate member countries (Australia, Canada, Japan, Korea, Mexico, New Zealand and the United States) and one Observer country (Morocco).

The ECMT is a forum in which Ministers responsible for transport, and more specifically the inland transport sector, can co-operate on policy. Within this forum, Ministers can openly discuss current problems and agree upon joint approaches aimed at improving the utilization and at ensuring the rational development of European transport systems of international importance.

At present, ECMT has a dual role. On one hand it helps to create an integrated transport system throughout the enlarged Europe that is economically efficient and meets environmental and safety standards. In order to achieve this, it is important for ECMT to help build a bridge between the European Union and the rest of the European continent at a political level.

On the other hand, ECMT's mission is also to develop reflections on long-term trends in the transport sector and to study the implications for the sector of increased globalisation. The activities in this regard have recently been reinforced by the setting up of a New Joint OECD/ECMT Transport Research Centre.

Publié en français sous le titre :
L' offre de transports : les limites de la (dé)réglementation

Further information about the ECMT is available on Internet at the following address:
www.cemt.org

TABLE OF CONTENTS

INTRODUCTORY REPORTS

DELIMITING REGULATORY NEEDS

Günter KNIEPS
Albert-Ludwigs University
Freiburg
Germany

TRANSPORT SERVICES: THE LIMITS OF (DE)REGULATION – ISBN 92-821-2345-6 - © ECMT, 2006

DELIMITING REGULATORY NEEDS

SUMMARY

Freiburg, February 2004

1. INTRODUCTION

Ever since the European Court of Justice ruled against the Council of Transport Ministers in 1985 for failing to ensure freedom to provide services in the sphere of international transport, the paradigm shift towards full competition on the European transport market has become irreversible. The European Union has played a leading role in this process, and the benefits of free access to transport markets throughout Europe are now largely uncontested.

At first sight the carriage of passengers and goods by road, rail, sea and air seems very heterogeneous. Common to all the different forms of transport, however, is the fact that access to transport services requires the use of transport infrastructure. This means the development and operation of not only transport infrastructure but also traffic control systems.

Unlike past studies, which tended to focus on one specific sector, this paper opts for a disaggregated approach which distinguishes between three network levels (Knieps, 1996):

(1) Transport services (carriage of passengers and goods by train, air, ship, lorry or private car);
(2) Traffic control systems (e.g. air traffic control, train control, road traffic control and information systems);
(3) Fixed infrastructure (e.g. railway tracks, stations, roads, airports).

With this approach it is possible to conduct a thorough analysis of competition potential and the continuing need for regulation in the future and, in particular, a separate study can be made of the question of access to transport infrastructure in connection with today's network economy.

Efficient competition on European transport markets is conditional upon the existence of non-discriminatory access to infrastructure for all active and potential transport service providers. In addition, however, efforts must also be made to ensure scant infrastructure capacities are shared out efficiently and costs are covered. This paper presents a disaggregated approach to regulation which enables these objectives to be met as comprehensively as possible.

Chapter 2 starts with an explanation of the theory of monopolistic bottlenecks, which can be used as a basis for distinguishing between parts of a network where competition functions efficiently and those which enjoy stable, network-specific market power. Chapter 3 explores the opportunities for potential and active competition on the markets for transport services, while Chapter 4 explains the potential for auction competition in respect of air traffic and train control systems. Chapter 5 is devoted to regulation in connection with access to transport infrastructure, including the relationship between disaggregated regulation of bottlenecks and the essential facilities theory, and the development of an appropriate approach towards regulation of access charges. Lastly, the paper looks at the ongoing reform of access to European airports (Chapter 6) and rail networks (Chapter 7).

2. THEORY OF MONOPOLISTIC BOTTLENECKS

An appropriate economic reference model, which exposes the need for action to control market power in network sectors, must be capable of grasping essential network characteristics (cluster/bundling effects, externalities, etc.) without automatically assuming these equate to market power. This chapter will attempt to show that stable network-specific market power and the ensuing need for regulation only exist in the event of monopolistic bottlenecks.

The theory of monopolistic bottlenecks is central to the disaggregated regulation approach in terms of locating network-specific market power in connection with the efforts to determine the minimum basis for regulation (cf. Knieps, 1997a, pp. 327 ff; Knieps, 1997b). The aim is to come up with a coherent basis, consistent with the network economy, which can be applied to all network sectors and which, regardless of historical or institutional quirks, provides justification for *ex ante* regulatory measures. The remaining network areas come under general competition law. In this context, the need for regulation is concerned in particular with the need to design a system for controlling access to monopolistic bottlenecks and for charging users. The problems associated with monopolistic bottlenecks and, in particular, the problem of network access (Baumol, Willig, 1999, p. 44; Knieps, 1997a, p. 327 ff.; Laffont, Tirole, 2000, p. 98) are currently frequent topics of discussion in the context of the network economy.

Network-specific market power can only be identified by consistently implementing Stigler's concept of market entry barriers. According to Stigler:

> *A barrier to entry may be defined as a cost of producing (at some or every rate of output) which must be borne by a firm which seeks to enter an industry but is not borne by firms already in the industry* (Stigler, 1968, p. 67).

Providing inputs are available to active and potential market players under the same conditions, Stigler considers there are no barriers to entry. Therefore, economies of scale, for example, do not constitute entry barriers providing newcomers to the market also have access to the same cost function. Stigler's concept also implies that traditional competition parameters, such as product differentiation coupled with the need to build up a good reputation and develop goodwill, or the capital required, are not entry barriers because they affect all active and potential enterprises equally. In other words, they are situations where the cost functions depend only on factors that are systematically available to all enterprises[1].

The conditions governing a monopolistic bottleneck are met when:

(1) a facility is essential for reaching customers, i.e. if no second or third such facility exists, in other words, if there is no *active* substitute. This is the case when cluster effects produce a natural monopoly and a single provider is able to make the facility available more cheaply than several providers[2];
(2) At the same time, the facility cannot reasonably be duplicated as a way of controlling the active provider, i.e. when there is no *potential* substitute. This is the case when the costs of

the facility are irreversible and there is therefore no second-hand market in operation for such facilities.

Consequently, network-specific market power in the hands of the established enterprise is only to be expected in part-areas, characterised by cluster effects and irreversible costs at one and the same time. Although they are no longer relevant for decisionmaking by established enterprises, as far as potential competitors are concerned, irreversible costs are a crucial factor, insofar as they must decide whether to invest such costs in the market or not. Established firms therefore have lower decision-relevant costs than their potential rivals. This means there is room for strategic manoeuvring, with the result that inefficient production or surplus profits no longer necessarily enable newcomers to enter the market. The market power of the firm which enjoys such a monopolistic bottleneck is therefore stable, even if all market players are fully informed, all users are prepared to switch to another provider, and small price adjustments have an effect on demand[3].

In the absence of irreversible costs, however, and as a result of the controlling effect of potential competition, cluster effects do not produce stable market power[4], regardless of the size of the relevant network operator's market share, insofar as inefficient providers of non-market-oriented services will be replaced by new entrants, owing to the pressure of competition. In this case, there is no need for regulation to limit the active operator's control over the market.

The bottleneck theory does not set out to deny the information problems encountered to varying degrees by real markets. *Ex ante* stable market power cannot be deduced from the existence of information problems, however, insofar as markets tend to be very good at (endogenously) developing institutions to overcome their information problems. Switching costs, which occur in many areas of the economy, are no explanation for monopolistic bottleneck situations either. Examples of switching costs include monthly or annual season tickets for concerts which cannot be transferred if the holder moves house, or the costs incurred by firms when employees leave as soon as they have "learnt the ropes", etc. They are no justification for regulatory measures and can be left to the market's own problem-solving ability (cf., for example, von Weizsäcker, 1984; Tirole, 1989, Chapter 8). However, the existence of network externalities is no justification for sector-specific regulation either. The essential feature of such externalities is that, for an individual, the advantage of being part of a network depends not only on its technical specifications – its standard – but also on how many others are involved in it. Where there are positive network externalities, the benefit for the individual increases with the number of other network members, in other words, the number of those using the same standard. In the absence of network-specific market power, negotiations between network operators can prove effective because both sides stand to benefit from the agreements. On the other hand, access to bottlenecks does present a need for regulation, given that network-specific market power allows for strategic manoeuvring, which also hampers full enjoyment of positive externalities associated with access to the network (cf., for example, Blankart, Knieps, 1995).

Indeed, one of the essential features of the ability of competition to operate on the free markets for transport services is that corporate strategies such as product and price differentiation, the build-up of goodwill and the development of an efficient distribution network, etc., can also be used for strategic purposes. Information problems (search costs, asymmetric information, etc.) can also play a role[5]. However, this must not lead to the opposite conclusion, namely, that competition basically does not work on transport markets, nor does it mean that general competition law should not be applied on these markets. What it does mean, however, is that, as on any other market characterised by organised competition, the burden of proof as to the existence of market power and as to whether such power is abused, rests with the competition authorities. In contrast to general *ex ante* regulation, such interference in competition should always be carried out only on a case-by-case and *ex post* basis[6].

3. COMPETITION POTENTIAL ON THE MARKETS FOR TRANSPORT SERVICES

Active and potential competition operates on the transport markets. The very fact that transport services are on offer in the form of a network with their associated bundling effects, implies that there is no monopolistic control where transport undertakings have free access to the market, since high profits recorded by one undertaking have the immediate effect of attracting others. There is no danger of preventing competitors from entering the market, insofar as the decision-related costs in respect of transport services are similar for established undertakings as for potential rivals. As a result, the irreversible costs associated with providing rail services on a railway network, for example, play no significant role. The use of trains is not confined to certain lines; they are just as mobile geographically as aeroplanes or lorries.

For competition to be effective, however, the conditions of access to the transport infrastructure must be the same for all (active and potential) service providers. If established undertakings have preferential access to scant infrastructure capacities, they enjoy unwarranted advantages over others, which can result in their gaining control over otherwise competitive markets.

Whereas the theory of contestable markets examines only the role of potential competition with identical cost functions for both active providers and potential rivals (cf. Baumol, 1982; Panzar, Willig, 1977), effective competition on the markets for transport services does not only mean potential competition Often a newcomer enters the market with no intention of duplicating the established undertaking. What is important is active competition achieved by means of technological and product differentiation, and the introduction of new products and processes. As a direct consequence of this, it is misleading to assume that newcomers have as their reference point the belief that ideally there can only be one transport network on the markets for transport services.

In the rail sector too active competition on busy lines should mean more efficient pricing, including more incentives for cost-efficiency and pressure to offer services tailored to meet demand. In the passenger sector, the pressure of competition reveals whether the length of the trains deployed and the intervals between them matches demand. Past supply concepts (e.g. clockface timetabling) are then brought into question when customers fail to honour them with a corresponding demand. Regular runs with (almost) empty "ghost trains" are no longer sustainable in a competitive context but, on the other hand, there are new incentives for providing a flexible supply of extra services at peak times. The entry of newcomers onto the market broadens the range of services offered extensively as well as widening consumers' choices in terms of price and service quality. Opportunities for new entrants include the detection and exploitation of gaps in the market, such as the development of a Europe-wide express service for passengers and goods, based on a high-performance, computer-assisted logistics system. Service improvements are also possible over shorter distances, however, with examples including a denser timetable offering better connections. In addition to the pressure of potential competition, active competitive between different undertakings is therefore also a source of potential which should not be underestimated.

4. COMPETITIVE POTENTIAL OF TRAFFIC CONTROL SYSTEMS

The provision of a transport service requires not only a vehicle (e.g. a train or aeroplane) but, at the same time, access to infrastructure (e.g. a railway line or landing slot). With rail and air transport, traffic movements must also be constantly monitored and co-ordinated. Train and air traffic control systems are needed, not only to guarantee traffic safety but also to allocate the available infrastructure capacity. Traffic control systems also look set to play an increasing role in the road sector.

It is important to remember that the provision of transport services requires simultaneous access to infrastructure and a traffic control system, regardless of whether these functions are vertically integrated within a single undertaking or whether they are performed by several different undertakings. Although airport operators, air carriers and air traffic control authorities can only guarantee smooth services by working together, they have always been separate from each other, both in terms of the way they are organised and institutionally. Such is not the case in the rail sector, however, where all functions used to be vertically integrated within the national railway companies and where co-operation between national companies tended to be minimal. But in this sector too there have been recent moves towards a disaggregated approach towards regulation. Competition on the railway networks is only possible if railway undertakings have unimpeded access to railway lines and at the same time can use the services provided by the different train control systems (cf., for example, Berndt, Kunz, 2003, p. 186 ff).

Train control systems are the crucial link between railway infrastructure and operations. Both the throughput of traffic and repairs carried out on the track must be co-ordinated by such systems. As in the air transport sector, the cost of such co-ordination is basically the same, regardless of whether only one or more than one railway undertaking is operating over the network. Rather, it depends on the number of trains and their operating speed.

Traffic control systems do not constitute monopolistic bottlenecks. They are natural monopolies, whose geographical limits have to be clearly defined (control jurisdiction). This still does not mean they enjoy network-specific market power, however, since the computer software and know-how needed to develop such systems are not tied to any particular place. Whereas with respect to transport services, the pressure of competition can also be achieved by selective, sequential (time-wise) hit-and-run entry (which does not necessarily result in total substitution of the established undertaking), in the case of traffic control systems auctioning competition is needed, where the subject of the auction is the predefined geographical traffic control area for a set period of time, and where the contract is awarded to the bidder who is able to offer a service at the lowest prices while at the same time covering the costs involved.

The move towards an EU infrastructure policy, with the aim of developing trans-European networks, automatically means interoperability must be promoted between the individual state networks[7]. There is a great need for co-ordination, particularly with regard to traffic control and monitoring systems. Clearly, active competition between different providers of traffic control systems cannot work. An individual aeroplane or train can only be monitored by one institution at a time, if chaos and accidents are to be avoided. Such systems are therefore natural monopolies with

geographical limits which have to be clearly defined at the level of institutions. Responsibility for traffic monitoring must remain in the hands of a single authority for a set period of time. This raises the question of the "natural" limit to a regional monitoring area and the co-ordination needed between different areas.

In the past, the railway monopoly led to a predominantly national approach to capacity management of railway lines and timetabling influenced by national considerations. International co-ordination and co-operation within the International Union of Railways was therefore minimal, both in terms of standardisation efforts and route management. Optimisation efforts were confined to the national rail systems (cf. Knieps, 1995). Since then, the tendency has been increasingly towards integration, including the development and introduction of a new, standard European Train Control System.

Like air traffic control systems, however, train control systems have considerable potential for operating across frontiers. Competition on the European rail transport markets and the ensuing rise in demand for European rail traffic mean there is a need for systematic internalisation of cross-border restrictions. For example, the technical limits of telecommunications and radio equipment should no longer coincide with countries' political borders. The full benefits of systems able to operate across borders need to be exploited so that competition can develop to its full potential on European rail transport markets.

The development of an integrated European train control system would be considerably facilitated by the establishment of independent agencies like those which exist for air traffic control. Insofar as no such integrated European system is being developed, steps should at least be taken to ensure full advantage of the opportunities which exist for intensive co-ordination and harmonisation of train control systems, for example, by stepping up standardisation efforts and co-ordination of timetabling.

Increased harmonisation of systems can be introduced in the context of a competition organised on an institutional level between national train control systems. If calls for tenders are issued throughout Europe (as is now generally the case with other services), train control agencies which are very successful in a given country are also likely to be successful with their auction bids in other countries. As a result, new train control developments in one country will gradually spread to other countries. In addition, competition between institutions will probably result in full utilisation of cost-saving opportunities and a better supply of transport services. The monopoly of national train control agencies on information is not as strong as it was. Transport undertakings now have an opportunity to put pressure on their respective agencies.

5. REGULATION OF ACCESS TO TRANSPORT INFRASTRUCTURE

On account of bundling effects, combined with irreversible costs, transport infrastructure (railway track, stations, airports, etc.) can result in the development of network-related market power. Even in open markets, the State has an important responsibility to regulate this monopolistic bottleneck. In particular, steps must be taken to ensure that there is no abuse of this market power in order to distort active and potential competition on complementary transport markets.

5.1. Disaggregated regulation of bottlenecks versus essential facilities doctrine

The concept of essential facilities plays a central role when competition rules are used to control network-specific market power. Facilities (infrastructure or equipment) are considered "essential" if they are simultaneously:

- indispensable for reaching customers and/or enabling competitors to do business;
- not present anywhere else on the market;
- such that they cannot objectively be rebuilt at a reasonable cost.

This concept is closely related to the essential facilities doctrine which derives from American anti-trust laws and is now increasingly applied in European competition law (cf., for example, Lipsky, Sidak, 1999; Haus, 2002; Aberle, Eisenkopf, 2002). The doctrine affirms that facilities can only be considered essential if the following two conditions are met: without access to the facilities, it is impossible in practice for competitors to enter the complementary market; and it is impossible for a provider on a complementary market to duplicate the facilities at a reasonable cost[8] and there is no substitute for them[9].

In the context of the disaggregated approach to regulation, the essential facilities doctrine is no longer applied on a case-by-case basis, as is usual in anti-trust laws, but instead to a category of cases, namely, monopolistic bottleneck facilities. The form taken by the conditions of non-discriminatory access to the essential facilities must be defined (cf. Knieps, 1997a; Knieps, 1997b). Insofar as monopolistic bottlenecks exist in liberalised network sectors, they call for specific residual regulatory measures to control the remaining market power. In particular, symmetrical access to the monopolistic bottlenecks must be guaranteed for all active and potential network service providers, so that competition stands a chance on all complementary markets.

The starting point of such a regulatory policy should be that regulatory measures are strictly confined to those network areas where market power potential actually exists. Regulation of access fees in respect of monopolistic bottlenecks must not therefore be accompanied at the same time by regulation of prices on the complementary markets for transport services (cf. Knieps, 2000, p. 100 ff). There are two further aspects which must be taken into consideration. First, it is wrong "automatically" to assume from competition on the service markets that there is no potential for market power on the infrastructure level, insofar as infrastructure fulfils the criteria governing monopolistic bottlenecks (cf. Brunekreeft, 2003, p. 89 ff). Second, there is the question of minimum regulation which, while being sufficient to guarantee non-discriminatory access to essential facilities, stops short of infringing excessively on the regulated enterprise's property rights[10].

5.2. Price-level regulation of infrastructure user charges

The effect of a total denial of access to infrastructure would also result if capacity was made available at prohibitively high prices. This alone shows that effective application of the essential facilities doctrine must be combined with appropriate regulation of access conditions. The identification of monopolistic bottlenecks is always based on an intramodal perspective, a decisive factor being the need for complementary service providers to have non-discriminatory access to such facilities. However, the existence of monopolistic bottleneck facilities does not necessarily guarantee that there will be long-term surplus profits. Firstly, there is the possibility of the "necessary case", where even unregulated train path providers are unable to meet their costs (Berndt, Kunz, 2003, p. 207 ff). Secondly, competition between modes can severely limit an infrastructure provider's profit potential (Fritsch, Wein, Ewers, 2003, p. 208).

Regulation of railway infrastructure access charges should always be confined to the parts of the network where market power potential actually exists. Price/profit regulation in the complementary, competitive parts of the network would go against the principle of minimalist regulation and lastingly obstruct the goals of a fully open market. Regulation of infrastructure user charges must not therefore lead, at the same time, to regulation of prices in the complementary parts of the network, where there is no market power potential.

Regulation of railway infrastructure access charges should be limited exclusively to price-capping. The basic principle underlying price-capping regulation is that price levels should be regulated in areas where there is network-specific market power. The benefits of price-capping in terms of efficiency improvements and future investment activities can only unfold if price-capping is applied in its "unadulterated" form and not combined with input-based profit regulation. Individual pricing agreements amount to over-regulation which is harmful to competition.

The reference point for monopolistic bottleneck facilities in the sense of quasi-competition, where the criticism of abuse of market power is not justified, should be overall cost recovery. Regulatory authorities should not force undertakings to apply specific price rules, such as Ramsey prices or two-part tariffs, as this would hamper their quest for innovative pricing systems. It is always possible that better rules will be found in future.

5.3. Flexible innovative pricing structures for network access

5.3.1 Advantages of the subsidiarity principle

Intramodal competition on European transport markets requires that all transport service providers, domestic or foreign, should have non-discriminatory access to infrastructure. The criterion of non-discrimination must refer here both to the quality of the available infrastructure (avoidance of grandfather rights, etc.) and the access tariffs.

Financing of transport infrastructure (roads, canals, airports and railways) used to be seen as a typical responsibility of the State. The provision of capacity utilisation is so low that there is no rivalry about use and market pricing based on load factors is not appropriate.

Unlike transport infrastructure where capacity is in short supply, parts for which demand is low have all the features of a public commodity because of the lack of rivalry. Accordingly, in the case of little-used transport infrastructure it is still necessary to decide on the level of investment which is socially (politically) desirable and to guarantee state funding. By themselves, however, high fixed costs for making transport infrastructure available are no grounds for unlimited state subsidies. Above all, the scale of subsidies must not be left to chance (Scientific Advisory Board at the Federal Ministry for Transport, Construction and Housing, 1999, p. 442). Rather, what is needed is transparent implementation of policies based on the "Orderer Principle", according to which lines operating at a loss, for example, may be funded via invitations to tender. More fundamental is the question of public investment policy, particularly with relation to federal infrastructure planning (Aberle, 2003, p. 453 ff).

One question which has to be asked is to what extent the goal of international intramodal competition calls for pan-European harmonisation of the pricing principles applied to infrastructure user charges. Providing the same conditions apply to all transport service providers in each country, there is no discrimination against foreign providers (cf. Scientific Advisory Board at the Federal Ministry for Transport, Construction and Housing, 1999, p. 443). On the other hand, different

infrastructure charges in different countries, *ceteris paribus,* lead to different transport tariffs, which in turn can encourage users to look for other solutions. It is to be expected, however, that the volume of traffic will also grow in international traffic as a result of the introduction of innovative, flexible pricing systems.

As far as possible, systems for charging for infrastructure access should ensure non-discrimination, the efficient allocation of scant infrastructure capacities (efficiency requirement) and a harmonised degree of cost recovery (financial requirement), all at the same time. Conventional full-cost calculations, based on administrative apportionment keys for assigning shared infrastructure costs to the different user groups, make no sense economically and are obviously not the solution (cf., for example, Baumol, Koehn, Willig, 1987). However, pricing based on marginal social costs cannot fulfil all these criteria at once either. In particular, fixed cost recovery is a residual value.

A critical feature, when it comes to making infrastructure capacity available, are the high fixed costs and the economies of scale which ensue with respect to service provision. Where there are economies of scale it is known that marginal cost prices no longer produce total cost efficiency. By themselves, the high fixed costs associated with providing access to network infrastructure are no grounds for state subsidies. Owing to the sharp rise in demand for transport in recent decades, infrastructure capacity is in increasingly short supply, creating a need to find free market solutions.

5.3.2 Welfare-enhancing price differentiation

Pricing principles should aim to incorporate the financing requirement as an *ex ante* condition, so that the State, which has to be brought in to mop up any shortfall, is not faced with any incalculable debt requirements. With the help of an appropriate two-part pricing system, however, consisting of a fixed use-related component (Infracard) and a variable component lower than a linear price, it is also possible to meet the underlying need for market-oriented infrastructure user charges. For this, inclusion of the demand side is vital. Tracking down pricing principles is therefore a business matter which ultimately can only be performed by the infrastructure companies themselves.

Under certain conditions, price differentiation can enable an undertaking to survive, but only if a standard (uniform) price fails to guarantee the necessary cost-efficiency[11]. For the national economy, a key advantage of two-part tariffs over one-part tariffs is that the goal of (partial) cost efficiency can be achieved without severely reducing transport demand by applying major mark-ups to the variable price so that scant infrastructure capacity can be allocated efficiently. Of particular importance for the welfare-enhancing effects is the increase in the volume of traffic (more traffic) typically associated with two-part tariffs[12]. Major infrastructure users (Infracard holders) can be expected to do whatever they can to make as much use of the infrastructure as possible. Moreover, small users for whom the purchase of an Infracard is not worthwhile are not excluded insofar as they also have access to infrastructure capacity.

Optional two-part tariffs have the advantage of persuading potential users to disclose information about their own particular willingness to pay (e.g. whether it is worth their while to pay a certain fixed entrance fee) and in so doing assign themselves to a particular user category.

It must be stressed here that there is no single system which is better than all the others and which could be centrally adopted as a pricing objective. Rather, it is a case of testing the limits of further price differentiation via a process of trial and error. The limit to further differentiation is when the transaction costs become too high for the system, in other words, when the cost of avoiding arbitrage outweighs the benefits of graduating prices further. This limit cannot be uniformly defined, however, as it will depend on the prevailing "local" circumstances. Consequently, there is a need for a

regulatory framework which does not impede infrastructure operators in their quest for new pricing structures.

6. CURRENT REFORM OF ACCESS TO EUROPEAN AIRPORTS

European air transport was liberalised over a decade ago. The third package of measures for the liberalisation of air transport in the European Union came into force on 1st January 1993, largely replacing the bilateral air transport agreements signed in the past between Member States and making it possible for EU nationals to establish air transport undertakings anywhere in the European Union. The package also provided for free access to all intra-Community routes and flexible fares for the services operating on those routes.

In January 1993, the Council of the European Communities also adopted a Regulation[13] on common rules for the allocation of slots at Community airports, which established a legally binding framework applicable in all Member States. The main features of the Regulation include the maintenance of "grandfather" rights, according to which the air carrier which has operated a slot in the previous scheduling period has priority over other air carriers in respect of that slot in the next scheduling period. Unlike the exchange of slots, no provision is made for slot trading or slot auctions (Niejahr, 1999).

For competition on European air transport markets to operate efficiently, however, non-discriminatory access to airports must be available to all active and potential suppliers of airline services. At the same time, efforts must be made to achieve efficient allocation of scant infrastructure capacities and to cover the costs involved.

The system used to allocate slots at busy airports in a way which distorts competition is the central concern in the ongoing air transport debate (Brunekreeft, Neuscheler, 2003, p. 254 ff), which is also focused on revision of EEC Regulation 95/93. In this context, those issues proving to be especially controversial are the abolition of "grandfather rights" and the feasibility of market-oriented approaches to slot allocation at saturated airports, in particular slot trading and slot auctions and the role of optimum user charges (Boyfield, 2003, p. 34 ff).

6.1. Slots as marketable commodities

It is an indisputable fact that many airports -- not only in the USA but also in Europe -- hit their capacity limits at peak times. In view of these capacity bottlenecks, which are getting worse, the public authorities are increasingly being asked to overcome the problem by developing more capacity. Investment on such a scale, however, would produce a surplus of slots which, economically speaking, would be a waste of valuable resources. This is not the same as saying that airport investments (generally) should stop, but rather that they should continue only for as long as the added benefits of capacity expansion are in keeping with the extra costs involved. This means that, even in the context of economically optimum investment, capacity at busy airports will still be in short supply at peak times. The allocation of scarce slots is therefore not only a transitional problem: the process of transforming a public commodity into a private one is unstoppable and irreversible.

The following two questions are examined in greater detail below:

- How can the allocation of scant airport capacities be organised in such a way that they can be put to optimum use from the point of view of air travellers (consumers)?;
- How can the allocation of such scant capacities be organised in such a way that distortions of competition on the liberalised markets for air transport can be prevented or kept to a minimum?

6.2. Economic characteristics of slots

As soon as airport capacity is no longer available in excess -- i.e. as public commodities -- it becomes necessary to specify and define which slots have become in short supply and when. In the usual world of trading in commodities (for example, grain), there is the microeconomic problem of defining the different categories of a given commodity (e.g. types of grain), but the precise time of the transaction is generally not crucial. The situation regarding airport capacities is completely different. There are a great many resources which must be co-ordinated with each other, time-wise, to the highest possible degree of accuracy. The very definition of a take-off or landing slot opens up a vast range of alternatives which can be crucial for potential transactions. If a take-off slot, for example, means only the right of a given airline to take off within a relatively long period of time, that right is worth much less than a guarantee that the airline can take off at a specific point in time without being subject to any delays. Some airlines, on the other hand, may prefer flexible operating times. Trading in slots therefore presupposes that take-off and landing rights have first of all been defined in a way which reflects both the needs of the airlines (and their passengers) and the operational and logistical possibilities of the airport operators.

In 1969, limits were imposed on the number of peak-time take-off and landing rights issued at five American airports (Chicago O'Hare, Washington National, New York Kennedy, La Guardia and Newark) (high-density rule). The right to take off or land during this period was referred to as a "slot", with slots lasting either half an hour or an hour. Slots were not classed as property and offered no guarantee of punctuality (Report of the Congress, 1995, p. 1 ff).

The definition of slots contained in Article 2 of EEC Regulation 95/93 also leaves considerable room for manoeuvre ("*the scheduled time of arrival or departure available or allocated to an aircraft movement on a specific date at an airport co-ordinated under the terms of this Regulation*"). Here too, it is *ex ante* co-ordination, with no guarantee of punctuality, no rules on priority and no means of enforcing the right to take off or land as a right of ownership.

This imprecise formulation of the right to use airport capacities is exactly what the airport operators want. There is no incentive for them to issue guarantees of punctuality (for specific flights) and accept the liability rules which would stem from such guarantees without, at the same time, benefiting from the scarcity rents. On the other hand, it is patently obvious that administrative management of capacities in short supply, with no financial incentives for all the parties involved, can produce a high degree of inefficiency.

It has become clear that slots cannot be defined independently of the market-economy instruments used in connection with the allocation system. Efficient slot allocation means making maximum use of airport capacities, while still complying with the relevant safety standards. Allocation systems which take account of individual air carriers' priority and punctuality preferences within a given slot period would have the advantage of making do on average with smaller

reserve/buffer zones. For example, it would be possible to apply a system based on different slot categories, where owners of expensive slots would have higher priority at take-off than cheap slot owners, who would sometimes have longer to wait.

6.3. Abolition of "grandfather rights" in favour of *ex ante* auctioning of take-off and landing slots

As in the past, rather than being reallocated according to changing needs, take-off and landing slots in Europe remain in the hands of the airline to which they were initially allocated, even if that airline does not use them or another airline would put them to better use. (*Ex ante*) flight schedule co-ordination is carried out by the airport co-ordinators appointed by each individual country[13b]. The exchange and transfer of slots is allowed in the context of mutual agreements between air carriers. Voluntary airline associations worldwide also negotiate flight schedule adjustments to take account of airport capacity limitations and to avoid unnecessary delays. Even if take-off times booked by airlines are in increasingly short supply, so far they have always been allocated free of charge. Economically, there is no justification for this unless there is sufficient capacity for all airlines to be able to take off and land at any one time. Otherwise, the airlines which are already well-established at a given airport have an asymmetrical competitive advantage over other airlines.

With *ex ante* auctioning of take-off and landing rights, the advantages of long-term flight scheduling could be maintained but, at the same time, the market would have to be opened up to newcomers (Wolf, 1995). The danger of stockpiling slots for the strategic purpose of gaining a competitive advantage over rival companies decreases the larger the air transport market in respect of which *ex ante* flight schedules are drawn up. Alternative routes and extensive product diversity create sufficient substitution options between different air carriers.

Given that, on expiry of the auction period, slots tend to return to the airport operator, the opportunities for air carriers to receive scarcity rents from the sale of slots is limited to trading in slots during an auction period. Insofar as airport operators are now party to the scarcity rents received from slots, scant airport capacities are allocated to the bidders who show the greatest willingness to pay. Income from the auctions can also be ploughed back into airport development projects, given that very high scarcity rents send out a signal to the economy that airport capacity is insufficient and that further airport development is required.

6.4. Potential for trading in slots

One of the essential features of Regulation 95/93 is that the Council does not challenge the existence of "grandfather rights", with the result that a slot which has been operated by an air carrier shall entitle that air carrier to claim the same slot in the next equivalent scheduling period [Art. 8, paragraph (1)a]. The Regulation also provides that carriers have an obligation to use 80 per cent of the slots allocated to them. With regard to slots allocated out of the slot pool, preference is given to new entrants, to whom 50 per cent of such slots must be allocated (Article 10, para. 7). According to the Regulation, slots may be freely exchanged between air carriers by mutual agreement. To date, however, there is no provision for the sale and purchase or leasing of slots.

Slots may only be exchanged by air carriers which already have a slot. The ban on compensation payments means that if no equivalent slots are available there is no incentive for an exchange, even if exchanging a slot would ultimately result in its being put to more appropriate use. It also encourages black-market trading in slots and circumvention of the trading ban. Efficient allocation would

therefore be better achieved if trading in slots were to be officially allowed. In terms of free competition, too, trading in slots is preferable to exchanges. New entrants to the market, however, always have the opportunity to buy slots. If slots are in very short supply, with the result that the economic value of slots used for a specific purpose (e.g. scheduled business flight) is very high, slot owners must accept a considerable loss of income if they decide not to sell. In other words, in an efficient slot trading system, the opportunity costs of using slots or selling them will be nearly the same.

There is also the question of to what extent slot trading favours hoarding and therefore hampers competition on the markets for air transport. Even if the declared aim of the rule that unused slots are returned to the pool is to reduce stockpiling ("use-it-or-lose rule"), the possibility of such hoarding cannot be totally excluded, *a priori*. Unlike straightforward swaps, however, the possibility of selling slots increases the opportunity costs of hoarding slots or using them for a less lucrative flight because of the scarcity rents which can be obtained from selling.

6.5. Optimum user charges based on scarcity rents

For as long as airport slots are allocated by applying "grandfather rights" rather than in auction procedures, it is unclear how far a reform of airport charges can achieve more efficient allocation of scarce capacity while at the same time improving the *status quo* in favour of symmetrical access conditions.

Up until now, airport charges have basically depended on the weight of the aircraft, their function being to help finance the airports, not to control the way available capacities are allocated. Aircraft weight and flight distance are no indication of a flight's (marginal) contribution to the shortage of capacity available to air traffic control authorities and airports, nor of the costs which ensue for all other transport players. The decisive factor in this respect is the demand at a particular time for airport and route capacities. In the short term, airport capacities are essentially unchangeable. In the event of unforeseen bottlenecks, the typical solution is for airports to ration capacity on a first-come, first-served basis.

Air carriers tend to ignore the constraints imposed on other aircraft and their passengers by an additional flight at a particular time (e.g. longer clearance times, longer delays and longer flight times). To take these constraints into account, one solution would be to levy a (time-based) congestion fee, equivalent to the congestion costs incurred by all other flights as a result of the extra flight. If demand for infrastructure capacity still exceeds supply, the solution would be to charge a market price which includes not only the congestion costs but also a scarcity rent. These are therefore capacity bottlenecks where there is direct rivalry in respect of take-off and landing slots.

The congestion charges or scarcity rents would need to be graduated according to the degree of capacity utilisation during a day and depending on the season, insofar as capacity utilisation for the same flight may vary. This would enable peak-time take-off and landing rights to be allocated more efficiently. Congestion charges operate like peak load prices but are not to be confused with them, since a (non-time-based) congestion charge would still have to be levied even if there was no change in capacity utilisation over the period and no fluctuation in the level of the congestion costs.

Another advantage of congestion charges with respect to the short-term allocation of slots is that when congestion charges are high during peak periods there is no incentive to hoard slots. Given that unless it can be proved that 80 per cent of the allocated slot sequences have been used[14], slots in

Europe are returned to the slot pool, the introduction of capacity-based congestion charges also reduces the negative effects of "grandfather rights".

6.6. Reform solutions with respect to airport take-off and landing fees

To date, only a few airports levy capacity-shortage-based take-off and landing fees. At London's Heathrow and Gatwick airports, landing fees based on peak load pricing have been charged since the early 1970s. In the morning and evening, a standard peak landing fee applies regardless of aircraft weight. The principle of peak load pricing is also applied, however, in relation to aircraft passenger and parking fees. Insofar as peak load times for aircraft landing, passenger clearance and parking are not the same, different peak periods are defined for each of these services, with different peak load prices.

Other airports have also introduced basic charges or minimum landing fees to deter smaller aircraft from using the airport at peak times. They include Toronto, Sydney and New York, as well as Frankfurt, Munich and Dusseldorf.

Airport charges based on capacity utilisation are a (partial) improvement on weight-based take-off and landing fees, although they cannot be compared with optimum congestion charges. Much resistance still has to be overcome on the part of the different parties involved[15].

6.7. Disaggregated regulation of airport market power

The question which now has to be asked is whether there is a danger that, with the introduction of slot auctions and scarcity rents, airport operators would exploit their monopoly position (at least regionally) and in certain circumstances even reduce the number of slots up for auction, with the result that slot prices would reflect not only scarcity rents but also the airport operator's market power. This fear cannot simply be brushed aside on the grounds that many airports are still state-owned enterprises which, in any case, serve to maximise public interests. Unlike the operation of flight services, airport infrastructure is bound up with irreversible costs. Once in operation, they cannot simply be transferred to another location in the way that an aircraft can. This means that the result of inefficiency or excessive airport charges will not be the construction of another airport, since two airports would not be able to survive together for long. The introduction of scarcity rents and the move away from the strict principle of cost efficiency confers a discretionary power on airport operators which needs to be regulated.

This is where the modern regulatory theory of price-capping can provide appropriate solutions. Regulation of airport market power is sometimes difficult to reconcile with user charges based on the scarcity of capacity. Since it is not possible to expand airport capacity in the short term, the levying of optimum user charges at airports where capacity is in persistently short supply gives rise to scarcity rents which are not necessarily compatible with a given regulatory restriction.

7. REFORM OF ACCESS TO THE EUROPEAN RAIL NETWORK

Even if the markets for rail transport were not initially in the forefront of the deregulation debate, competition on these markets is now also seen as a central co-ordination instrument. Of particular importance in this context was the package of Directives, adopted on 26 February 2001 by the European Parliament and the Council, on the allocation of railway infrastructure capacity and the charging of infrastructure fees[16], building on the earlier Directives adopted in 1991[17] and 1995[18].

The provision of rail services requires not only trains but also access to a network of railway lines. Train control systems are also needed, not only to ensure traffic safety (traffic control) but also to ensure real-time train path management. Efficient competition on the markets for rail transport is conditional upon non-discriminatory access to the rail network for all active and potential providers of railway services.

7.1. Non-discriminatory access to railway infrastructure

Railway infrastructure (unlike rail services) is characterised by a monopolistic bottleneck situation, insofar as infrastructure operators have a natural monopoly and the construction of railway lines involves irreversible costs.

Insofar as monopolistic bottlenecks exist in network sectors, there is a need for specific residual regulation to control the remaining market power. In particular, all active and potential providers of network services must be guaranteed symmetrical access to the monopolistic bottlenecks, so that competition can operate properly on all complementary markets.

7.2. Efficient use of train paths based on scarcity prices

In the past, railway infrastructure capacity was traditionally allocated with the help of administrative measures defined by the national railway monopolies (e.g. timetable conferences, priority rules for determining train sequences in the event of delays and discretionary, one-off measures with respect to train control). There were no user charges to reflect the scarcity of capacity, even if at certain times of the day or year there were severe bottlenecks on certain sections of line. The decision when to use a particular line section therefore had no impact on fares, with the result that there was no incentive for peak-time customers to switch to less busy times. Customers who lay great store by punctuality and were willing to pay for it had no possibility of travelling in trains guaranteed to arrive on time.

One way of overcoming this problem is to levy a (time-based) scarcity price for using the railway network. The right to operate trains on busy sections of line could be auctioned among the different operators. Undertakings wishing to operate a train on a busy section would then have to pay a market price which reflects the opportunity costs of using these capacities. It is thereby possible that a freight train with traditionally low priority might be prepared to pay more than a traditionally higher priority

intercity train, to ensure that certain just-in-time production processes do not come to a standstill. If railway operators then pass these efficient access fees onto the user, customers must be required to pay higher fares at peak times and lower fares at less busy times. The function of this peak load pricing system is to manage available capacity.

7.3. Price differentiation versus discrimination

Even on busy routes, optimum user charges still may not enable all costs to be recovered. Economies of scale with respect to the construction of railway infrastructure are such that optimum access charges are unable to cover the infrastructure investment costs. This raises the question of how to finance the shortfall as well as the related question of the *ex ante* politically determined degree of cost recovery. To ensure that the incentives for achieving the necessary cost recovery are credible for the infrastructure operator, the degree of total cost recovery must not be left to chance (*ex post*), nor must it therefore be subject to constant scrutiny in terms of its level and validity.

The goal of efficient allocation of train path capacities in a context where cost recovery is limited calls for price differentiation strategies. In terms of access to train paths of different qualities, price differentiation has to take account of variations in the load factor, as reflected in differing degrees of willingness to pay (price elasticity of demand for rail infrastructure capacities). In particular, this means that differences in train path prices are due not only to variations in the cost of the diverse qualities of train path available but also to the differing additional price components which have to be included to cover the fixed infrastructure costs (cf. Berndt, Kunz, 2003, p. 195 ff).

In addition to a variable use-based component, there can also be a fixed pricing component (Infracard)[19]. The legal concept of discrimination must not be used to prohibit economically desirable price differentiation as such. This can be a danger, not least because another name in English for the economic concept of price differentiation, which is neutral in terms of competition, is "price-discrimination". Price differentiation means that differences in price are not solely due to differences in the costs directly attributable to a given customer but also take account of demand factors.

NOTES

1. On the other hand, the different entry barriers found, according to Bain (1956), in the traditional industrial economy (economies of scale, product differentiation, high capital needs, etc.) are not reliable proof of stable market power (cf., for example, B. Schmalensee, 1989). Von Weizsäcker (1980a; 1980b) shows, for example, that reputation and goodwill are effective ways of reducing insecurity, which can enhance social well-being. According to Stigler, the development of goodwill is not a barrier to market entry because it does not result in cost asymmetries between established firms and newcomers to the market.

2. A natural monopoly exists when the cost function in the relevant sector is subadditive in relation to demand. In the case of single products, economies of scale are sufficient for there to be a natural monopoly. In the context of a study of the cost factor in the case of multiple products, cluster effects stand out, owing to the economies of scale and bundling effects associated with the provision of services. As a result of these cluster effects, it may be possible for a single network provider to serve a given region more cheaply than several providers, and thus to enjoy a natural monopoly (cf., for example, Baumol, 1977).

3. This is Bertrand Nash's behavioural assumption, based on the theory of contestable markets (cf., for example, Baumol, Panzar, Willig, 1982).

4. In the absence of irreversible costs, there is no evidence in the case of a natural monopoly of market power, capable of withstanding alternative behavioural assumptions (cf. Knieps, Vogelsang, 1982). Market power based on the Cournot-Nash assumption becomes immediately unstable with the switch to the Bertrand-Nash behavioural assumption. Action taken by competition authorities would therefore have to refer to behavioural assumptions which are difficult to verify in practice.

5. The Bertrand Nash assumption, based on the contestable markets theory, does not set out to deny the information problems encountered on real markets either. Stable market power cannot be deduced simply from the existence of information problems, insofar as markets tend to be very good at (endogenously) developing institutions to overcome their information problems, for example, by building up goodwill. Conversely, stable market power is also to be found in natural monopolies with irreversible costs when all market players have all the information they need.

6. In this context, the competition authorities must weigh up two potential sources of error. Firstly, false positives can occur when the authorities interfere in the competition process, even though competition is working and there is no need at all for action in terms of competition policy. Secondly, false negatives occur when the competition authorities fail to act, even though competition policy calls for action (cf. Knieps, 1997c, p. 51).

7. Cf. Treaty of Amsterdam, Title XV, Trans-European Networks, Art. 154(2).

8. It is not possible, for example, to operate a ferry service without access to ports.

9. For a summary, see Areeda, Hovenkamp (1988). A further criterion for the essential facilities doctrine is sometimes that shared use of the facilities is essential for competition on the complementary market because it lowers the prices there or increases the volume of services offered. However, this criterion describes only the effects of access.

10. In principle, a distinction must be made between the question of whether network-specific market power exists as a result of a monopolistic bottleneck and the question of what constitutes appropriate regulatory measures.

11. If the average cost curve lies above the demand curve, price differentiation is unavoidable for full cost efficiency.

12. It should be borne in mind that environmental costs have to be met by means of separate environment policy measures (mineral oil tax, etc.).

13. Cf. Council Regulation (EEC) No. 95/93 of 18 January 1993 on common rules for the allocation of slots at Community airports, in: *Official Journal*, L 014 of 22 January 1993.

13b) Cf. aforementioned Council Regulation (EEC) No. 95/93, particularly Articles 8 and 10.

14. Cf. Council Regulation (EEC) No. 95/93, Article 10, paragraph 5.

15. In 1998, a reform of charges levied at Boston airport, with a view to significantly raising landing fees for smaller aircraft in the interests of more efficient capacity utilisation, met with such fierce resistance that it had to be abandoned.

16. Directives 2001/12/EC, 2001/13/EC, 2001/14/EC.

17. Directive 91/440/EC.

18. Directive 95/19/EC.

19. Two-part tariffs were levied by Deutsche Bahn AG in the context of the 1998 route pricing system (cf. Knieps, 1998).

BIBLIOGRAPHY

Areeda, P., H. Hovenkamp (1988), "Essential facility" doctrine? Applications, in: Areeda, P., H. Hovenkamp (eds.), *Antitrust Law*, 202.3 (Suppl. 1988), 675-701.

Aberle, G. (2003), Transportwirtschaft. Einzelwirtschaftliche und Gesamtwirtschaftliche Grundlagen, München, 4th edition.

Aberle, G., A. Eisenkopf (2002), Schienenverkehr und Netzzugang. *Giessener Studien zur Transportwirtschaft und Kommunikation*, Vol. 18.

Bain, J.S. (1956), *Barriers to New Competition*, Cambridge, MA, Harvard University Press.

Baumol, W.J. (1977), On the Proper Cost Test for Natural Monopolies in a Multiproduct Industry, *American Economic Review*, 67, 809-822.

Baumol, W.J. (1982), Contestable Markets: An Uprising in the Theory of Industry Structure, *American Economic Review*, 72, 1-15.

Baumol, W.J., M.F. Koehn, R.D. Willig (1987), How Arbitrary is "Arbitrary"? -- or, Toward the Deserved Demise of Full Cost Allocation, *Public Utilities Fortnightly*, 3, 16-21.

Baumol, W.J., J.C. Panzar, R.D. Willig (1982), *Contestable Markets and the Theory of Industry Structure*, New York.

Baumol, W.J., R.D. Willig (1999), Competitive Rail Regulation Rules -- Should Price Ceilings Constrain Final Products or Inputs?, *Journal of Transport Economics and Policy*, 33/1, 43-54.

Berndt, A., M. Kunz (2003), Immer öfter ab und an? Aktuelle Entwicklungen im Bahnsektor, in: Knieps, G., G. Brunekreeft (eds.), *Zwischen Regulierung und Wettbewerb -- Netzsektoren in Deutschland*, Heidelberg, 2nd edition, 165-218.

Blankart, Ch.B., G. Knieps (1995), Market-oriented Open Network Provision, *Information Economics and Policy*, 7, 283-296.

Boyfield, K. (2003), Who owns airport slots? A market solution to a deepening dilemma, in: Boyfield, K. (ed.), *A Market in Airport Slots*, The Institute of Economic Affairs, London.

Brunekreeft, G. (2003), Regulation and Competition Policy in the Electricity Market -- Economic Analysis and German Experience, *Freiburger Studien zur Netzökonomie,* 9, Baden-Baden.

Brunekreeft, G., T. Neuscheler (2003), Preisregulierung von Flughäfen, in: Knieps, G., G. Brunekreeft (eds.), *Zwischen Regulierung und Wettbewerb -- Netzsektoren in Deutschland*, Heidelberg, 2nd edition, 251-280.

Fritsch, M., T. Wein, H.-J. Ewers (2003), *Marktversagen und Wirtschaftspolitik*, München, 5th edition.

Haus, F.C. (2002), *Zugang zu Netzen und Infrastruktureinrichtungen*, Köln *et al.*

Knieps, G. (1995), Standardization: The Evolution of Institutions *versus* Government Intervention, in: Gerken, L. (ed.), *Competition Among Institutions*, London, 283-296.

Knieps, G. (1996), *Wettbewerb in Netzen -- Reformpotentiale in den Sektoren Eisenbahn und Luftverkehr*, Tübingen.

Knieps, G. (1997a), Phasing Out Sector-Specific Regulation in Competitive Telecommunications, *Kyklos*, 50/3, 325-339.

Knieps, G. (1997b), The Concept of Open Network Provision in Large Technical Systems, *EURAS Yearbook of Standardization*, 1, 371-390.

Knieps, G. (1997c), Wettbewerbspolitik, in: Börsch-Supan, A., J. von Hagen, P.J.J. Welfens, *Springers Handbuch der Volkswirtschaftslehre*, Berlin *et al*., 39-79.

Knieps, G. (1998), Das neue Trassenpreissystem: Volkswirtschaftliche Vorteile eines zweistufigen Systems, *Internationales Verkehrswesen*, 50/10, 466-470.

Knieps, G. (2000), Interconnection and Network Access, *Fordham International Law Journal*, 23, Symposium, 90-115.

Knieps, G., I. Vogelsang (1982), The Sustainability Concept under Alternative Behavioural Assumptions, *Bell Journal of Economics*, 13/1, 234-241.

Laffont, J.-J., J. Tirole (2000), *Competition in Telecommunications*, The MIT Press, Cambridge, Massachusetts, London.

Lipsky, A.B., J.G. Sidak (1999), Essential Facilities, *Stanford Law Review*, 51, 1187-1249.

Niejahr, M. (1999), Europäische Aspekte des Zugangs zu Infrastrukturen: Das Beispiel der Flughäfen, in: *Diskriminierungsfreier Zugang zu (Verkehrs-)Infrastrukturen: Konzepte, Erfahrungen und institutionelles Design*, Schriftenreihe der Deutschen Verkehrswissenschaftlichen Gesellschaft e.V., Schriftenreihe B, Vol. 224.

Panzar, J.C., R.D. Willig (1977), Free Entry and the Sustainability of Natural Monopoly, *Bell Journal of Economics*, 8, 1-22.

Report of the Congress (1995), *A Study of the High Density Rule*, May 1995.

Schmalensee, R. (1989), Inter-Industry Studies of Structure and Performances, in: Schmalensee, R., R. Willig (eds.), *Handbook of Industrial Organization*, Amsterdam *et al*., 951-1009.

Stigler, G.J. (1968), Barriers to Entry, Economies of Scale and Firm Size, in: Stigler, G.J. (ed.), *The Organization of Industry*, Homewood, Ill., Irwin, 165-170.

Tirole, J. (1989), *The Theory of Industrial Organization*, 2nd Printing, Cambridge *et al.*

Weizsäcker, C.C. von (1980a), A Welfare Analysis of Barriers to Entry, *Bell Journal of Economics*, 11, 399-420.

Weizsäcker, C.C. von (1980b), *Barriers to Entry: A Theoretical Treatment, Lecture Notes in Economics and Mathematical Systems*, Berlin *et al.*

Weizsäcker, C.C. von (1984), The Costs of Substitution, *Econometrica*, 52/5, 1085-1116.

Wissenschaftlicher Beirat beim Bundesminister für Verkehr, Bau- und Wohnungswesen (1999), Faire Preise für die Infrastrukturbenutzung, *Internationales Verkehrswesen*, 51, 436-446.

Wolf, H. (1995), Möglichkeiten und Grenzen marktwirtschaftlicher Verfahren zur Vergabe von Start-Landerechten auf Flughäfen -- Vorschlag für ein „zweitbestes Auktionsverfahren“, Kiel Working Document No. 671, Institut für Weltwirtschaft, Kiel.

WHERE DO WE STAND ON TRANSPORT INFRASTRUCTURE DEREGULATION AND PUBLIC-PRIVATE PARTNERSHIP?

Antonio ESTACHE
The World Bank
Washington DC
United States
ECARES
Université Libre de Bruxelles
Belgium

Tomas SEREBRISKY
The World Bank
Washington DC
United States

WHERE DO WE STAND ON TRANSPORT INFRASTRUCTURE DEREGULATION AND PUBLIC-PRIVATE PARTNERSHIP?

SUMMARY

Washington, February 2004

The findings, interpretations and conclusions expressed in this paper are entirely those of the authors and do not necessarily represent the official view of The World Bank.

1. INTRODUCTION

The history of the transport sector shows that, whether the focus is on buses, railways, roads or even airports and ports, the tale of public-private partnerships and their regulation is one of recurring policy themes. This recurrence is what allows us to draw better lessons over time as experiences evolve — and supposedly allows their internalisation in the next wave of policy reforms. The learning is slow because there are many dimensions to address, sometimes including dimensions complex enough to make the deregulation and ownership issues more of a significant anecdote rather than the centre of the debate. The history of the railways sector in the UK nicely motivates the associated policy, competition and regulation issues that arise in the context of the recurring cycles which characterise the evolving role of the State in transport[1].

The first railway carrying passengers and freight in Great Britain, the Stockton & Darlington Railway, was created in 1825 as a private company, much more willing to take the chance on a new vision than was the public sector. The State had no or little role and no costs. The network was developed through a piecemeal approach in which each line was being promoted by a small, private company, relying on the equivalent of Build-Operate-Own (BOO) projects for the times. The system was thus efficient in that it was following demand but inefficient in that it did not reflect the potential economies of scale. Large, long-distance operators eventually took over the smaller lines, improving the economies of scale. Eventually, in 1923, almost all the railways in the country were grouped into four large, new companies. The State started to become worried by concerns for competition in the sector. As a result of a drop in demand in the late 1940s, the sector became financially unviable and the State took over through a nationalisation of all assets. Throughout the 1950s, the sector continued to be restructured with a view to shrinking it, adjusting supply to demand to cut fiscal costs. The 1970s saw mixed modernisation efforts managed by the public sector at increasing fiscal costs. The 1980s, as a result of yet another macroeconomic shock -- the 1979 oil price shock -- saw a major deterioration in the demand for railways services and a further fiscal hit. By the 1990s, steps were put in place to restructure (including deregulating and privatisating) the railways to cut the fiscal burden. The circle seemed to have closed but not quite. By the end of the decade, unhappiness with the private model grew, to a large extent because of concerns for safety, while the trains were at the same time becoming fashionable again, largely due to a growing concern for congestion and pollution. The Government is thus now returning to the sector, as a very active regulator and possibly financier, and a new hybrid model seems to be emerging, starting what is likely to be a second turn on the enlarged "ownership circle".

The main purpose of this paper is to show that, to a large extent, the evolution of public-private partnerships in developed and developing countries since the early 1990s is very similar to the one observed for the British railways: private initiatives work for some things for a while, then some internal but often major shock external to the sector takes place and the public sector comes into the picture, first as a regulator then as an owner or at least financier. After some time, the public sector runs itself into problems and tries to get the private sector back. Eventually, hybrid solutions are found to ensure the survival of a sector for which the demand is strong, the economic impact brutal, but for which the financing structure needs to better account not only for the financial cost of the business but also for the major economic, social and political dimensions.[2] After a decade of increased

private sector participation, the cycle is now at the stage of its evolution at which governments need to define how hybrid they want their transport systems to be, as well as the specific, actual responsibility they want to assign to themselves.

To document this story and to show that we are at a point in history at which the cycle is likely to take us back to an increased role for the public sector in transport infrastructure, the paper is organised as follows. In Chapter 2, we provide a sense of the absolute and relative importance of the private sector in transport infrastructure from a global and regional viewpoint. In Chapter 3, we give a sense of the main types of contractual arrangements which have been adopted and the associated problems. In Chapter 4, we review the evidence on the effect of deregulation and re-regulation on the relative efficiency of public and private provision of transport services. In Chapter 5, we discuss the main new regulatory and competition issues emerging in the sector and their social consequences. Chapter 6 concludes.

2. HOW BIG IS THE ROLE FOR THE PRIVATE SECTOR IN TRANSPORT INFRASTRUCTURE?

The specific measurement of deregulation in transport infrastructure is not an easy task. Most of the information is anecdotal and, unlike what is happening in telecoms or energy for instance, the international associations for the sector do not keep track of the progress made by following a standard typology of reforms. Because most of the information available is anecdotal, it can hardly be organised into a detailed, useful typology. The big picture is, however, relatively straightforward and is best illustrated by the evolution of reforms in rail and ports.

Railways have now been or are being deregulated in most countries of the world, although with very diverse degrees of commitment to competition in and for the market. In most of North, Central and South America, in the largest countries of Oceania, in some of the major systems of Africa and in many of the largest countries of East Asia (with the noticeable exception of China), the private sector is now the key player, with the public sector confined to a regulatory role. This varies from a simple role as contract supervisor to the co-ordinator of complex timetables, the monitor of service quality and the key actor in tariff revisions when these are needed. Europe is also on its way to deregulating its railways but following a very "European" model, unmatched in other parts of the world. The main region which has opted out so far is Southern Asia.

Ports have also enjoyed strong reforms. Since the early 1990s, port deregulation has been widespread and the private sector is now one, if not the, key actor in most large countries of America, Africa, Oceania and even in the largest countries of Asia, such as China and India. As in the case of railways, Europe lags behind the other regions in terms of deregulation and continues to have a hybrid organisational structure in which the public sector still plays a key role as provider, financier and regulator, simultaneously. The overall picture of the sector can be summarised with some estimates made by Drewry (2002). In 1991, the public sector handled about 42 per cent of container port throughput, but by 2001 its share had dropped to about 27 per cent, with a strong concentration in South East Asia and Western and Eastern Europe.

One indicator, which gives a sense of the outcome of these restructuring and deregulation efforts over the last eighteen years, is the number of new transport infrastructure projects considered by the private sector as countries were liberalising the sector. Public Works Financing roughly compiles this information. According to this source, between 1985 and 2003, 1 137 new transport projects, worth US$684 billion, were planned around the world. This implies an average project size of US$600 million. The distribution of these projects by transport subsector shows that toll roads account for 50 per cent while rail, airports and ports account for 22, 16 and 12 per cent of the total commitments, respectively. Their distribution across regions shows that about 60 per cent of these projects are committed to developing countries but 55 per cent of the amounts committed are for developed countries. Table 1 also shows that the average project size varies significantly across regions, with the largest average projects in Europe and Asia. The average project size in developing countries is about half the size of that in developed countries.

While these commitment numbers give a useful sense of the upper limit of the private sector flows in the sector, they may be somewhat misleading. Indeed, only half of the new projects reported up to the end of 2003 had effective financing -- i.e. were under construction or operation[3]. The evolution of these commitments is interesting, however. It shows that, since the 1997 Asia crisis, the number of commitments has dropped significantly but much more importantly in developing than in developed countries.

Overall, this implies that, even though the deregulation of transport did generate enthusiasm during the early part of the 1990s, the Asia crisis and the overall instability of financial markets were enough to slow down and even overturn these commitments by the private sector. Many more reasons contribute towards explaining why effective disbursement lags behind. Some of them are country specific -- pace of the privatisation/liberalisation process, institutional organisation, rule of law -- while others are influenced by global conditions, mainly crises in financial markets. A more detailed look at the differences between developed and developing economies can provide useful insights.

Table 1. **Evolution of total planned commitments to transport projects in the World**
Investment commitment amounts, numbers of project commitments and average size of project commitments

	Total	Toll Roads			Rail			Airports			Seaports		
	Cumulative												
	1985-2003	1985-95	1995-98	1998-03	1985-95	1995-98	1998-03	1985-95	1995-98	1998-03	1985-95	1995-98	1998-03
Number of projects committed													
North America	**260**	**111**	***-17***	51	**16**	***3***	41	**15**	***17***	18	**15**	***-6***	-4
Europe	**297**	**67**	***28***	76	**34**	***48***	-21	**9**	***17***	19	**3**	***6***	11
Asia	**319**	**120**	***16***	-14	**96**	***18***	-30	**26**	***34***	-2	**30**	***13***	12
Other	**261**	**83**	***67***	-15	**27**	***26***	-10	**17**	***17***	-9	**21**	***27***	10
Total	**1137**	**381**	***94***	98	**173**	***95***	-20	**67**	***85***	26	**69**	***40***	29
Developed	**451**	**116**	***3***	135	**120**	***-24***	18		***24***	19	**3**	***3***	14
Developing	**686**	**265**	***91***	-37	**53**	***119***	38	**47**	***61***	7	**63**	***37***	15
Total commitment amount (in US$ billion)													
North America	**111.5**	**34**	***-8.1***	31.2	**21.6**	***12.1***	11.5	**4.7**	***0.2***	2.9	**0.85**	***0.97***	-0.47
Europe	**240.1**	**53.1**	***19.5***	34.6	**47.4**	***38.2***	30.2	**6**	***9.7***	-1.09	**0.08**	***5.06***	-2.7
Asia	**250.7**	**83.6**	***12***	-21.6	**54.6**	***26.1***	15.3	**51.8**	***25.2***	-20.6	**19.8**	***-5.4***	9.9
Other	**81.6**	**19.4**	***10.9***	-9.8	**12.8**	***11.1***	13.4	**3**	***3.5***	7.09	**2.07**	***3.07***	5.07
Total	**683.8**	**190.1**	***34.3***	34.4	**136.4**	***87.5***	70.4	**65.5**	***38.6***	-11.6	**22.8**	***3.7***	11.8
Developed	**379.7**	**85.6**	***4.8***	59.4	**69.3**	***39.9***	81.7	**41.2**	***7.1***	-11.7	**.1**	***1.3***	.9
Developing	**304.1**	**104.5**	***29.5***	-56	**67.1**	***47.6***	-11.3	**24.3**	***31.5***	0.1	**22.7**	***2.4***	10.9
Average project size (in US$ million)													
North America	**429**	**306**	***476***	612	**1 350**	***4 033***	280	**313**	***12***	161	**57**	***(162)***	118
Europe	**808**	**793**	***696***	455	**1 394**	***796***	(1 438)	**667**	***571***	(57)	**27**	***843***	(245)
Asia	**786**	**697**	***750***	1 543	**569**	***1 450***	(510)	**1 992**	***741***	10 300	**660**	***(415)***	825
Other	**313**	**234**	***163***	653	**474**	***427***	(1 340)	**176**	***206***	(788)	**99**	***114***	507
Total	**601**	**499**	***365***	351	**788**	***921***	(3 520)	**978**	***454***	(450)	**330**	***93***	407
Developed	**842**	**738**	***1 600***	440	**578**	***-1 663***	4 539	**2060**	***296***	-616	**33**	***433***	64
Developing	**443**	**394**	***324***	1 514	**1 266**	***400***	-297	**517**	***516***	14	**360**	***65***	727

Source: Public Works (2003) and authors' calculations.

2.1. Developed economies

Two-thirds of the private projects credited to developed countries accrue to five countries: the US (122 projects), Spain (66), UK (64), Australia (46) and Canada (37). Adding Portugal (22), Germany (19), Ireland (16) and Greece (14) covers 90 per cent of the projects undertaken during the period 1985-2003. In terms of commitments, the ranking is somewhat different. The UK enjoys 27 per cent of the commitments, the US 17 per cent, Japan 8 per cent, Italy and Spain 7 per cent, Canada 6 per cent and Australia 5 per cent. The sectoral composition of the projects reveals nicely the subsectoral demand. Toll roads and rail projects attract the most attention, reflecting the increased demand for passenger mobility in developed countries and in Europe due to the increased integration of the continent. Seaports do not seem to need new investments, which is consistent with the wide perception that there may be an excess capacity in the sector and that many of the ports continue to be maintained for political or strategic reasons. Airport projects are heavily concentrated in the United States, Australia and the UK, where the size of the first two countries and the insular character of the third ensure that the increased demand for passenger mobility is not monopolised by ground transportation.

The statistics also reflect the levels and commitments of the governments of the top three countries to private sector participation in the transport sector. The US has indeed a long tradition, which for the last fifteen years or so has been actively followed by the UK and Australia. More recently, following the liberalisation stimulus of the European Commission, many countries in continental Europe are actively involving the private sector in the transport sector. In particular, Portugal, Greece and Spain have experienced a significant increase in the quantity of projects planned, often with the financing of the regional development funds of the EEC.

2.2. Developing economies

The commitments published by Public Works are known to be overestimates of the actual numbers but also to sometimes fail to account for new projects in some regions and sectors in developing countries. For some time now, the World Bank has been publishing more refined estimates of these commitment figures, trying to separate wishful thinking by sponsors from serious commitments. This database covers actual deals -- rather than planned projects, as in the case of the Public Works database -- on projects which have reached financial closure in water, electricity, natural gas, telecommunications and transport. It deals exclusively with projects in which the private company assumes operating risk during the operating period or assumes development and operating risk during the contract period. Investments and privatisation revenues are recorded on a commitment basis in the year of financial closure. Actual disbursements are not tracked.

This effort results in a significant drop in the estimates to US$135 billion between 1990 and 2001, implying that the total commitments to transport in the world were at most US$515 billion rather than US$684 billion. For developing countries, this implies an average project size of US$204 million, or less than 50 per cent of the estimate provided in Table 1. Considering that the world counts 155 countries defined as developing, the average commitment per country would be about US$80 million per year[4]. The average project size implies that at least half of the countries do not benefit from any project. The privatisation experience has, in fact, been even more concentrated than expected. Indeed, China, Brazil, Argentina, Mexico and Malaysia accounted for most of the investment during the period, although with strong fluctuations over time. Up to the Tequila Crisis, these countries accounted for more than 80 per cent of total investments, reducing their share to 50 per cent in 1995. In 1998, their share was again over 80 per cent, but after the Asian crisis it has been falling, reaching just over 30 per cent in 2002. Globally, the developing countries' transport sector

only attracted about 18 per cent of the total investment flows which accrued to the infrastructure sector. It lags significantly behind the telecommunications (44 per cent) and electricity (28 per cent) sectors.

Table 2. **Estimates of transport infrastructure projects having reached financial closure in developing and transition economies (1990-2001)**

	Africa	East Asia	Eastern Europe	Latin America	Middle East	South Asia	Total
AIRPORT							
Number of transactions	10	17	14	32	7	2	82
Value in millions of US$	375	2 625	1 500	7 125	750	125	12 500
PORT							
Number of transactions	10	52	16	78	8	13	177
Value in millions of US$	180	8 820	720	5 580	900	1 800	18 000
RAIL							
Number of transactions	9	11	7	48	1	0	76
Value in millions of US$	288	10 080	288	17 856	288	0	28 800
ROAD							
Number of transactions	9	149	6	137	0	26	327
Value in millions of US$	2 280	34 200	2 770	36 000	0	750	76 000
TOTAL							
Number of transactions	***38***	***229***	***43***	***295***	***16***	***41***	***662***
Value in millions of US$	***3 123***	***55 725***	***5 278***	***66 561***	***1 938***	***2 675***	***135 300***

Source: World Bank, Infrastructure Vice-Presidency, PPI Database.

The sectoral distribution of projects (measured in US dollars) is just as skewed toward road deals as it is in developed countries, simply because the increase in demand for passenger mobility is also very high. The distribution of projects in the developing world also shows, however, the impact of trade liberalisation and of globalisation and their catching up with regional integration, long established within the US and across European countries. Freight needs to be moved and exported. Ports are a key element in that strategy. In Latin America, this concern implies that railways also need to be scaled up, because in many countries declining reliability and overall quality was such that, except for captive shipments, the rail sector had been losing a large share of its potential market to the trucking industry. In Brazil, for instance, the odds of relying on rail rather than trucks were lower the longer the delivery distance! In general, it is becoming increasingly clear to governments that both efficient logistics and the opportunity for multimodal arrangements are essential to competitiveness. The most perceivable outcome of the need to reduce logistic costs is that the private sector will be responsible for large shares of transport services. It is not too risky to predict that this trend is likely to persist as the demand for transport services continues to grow much faster than the government's ability to raise the resources to finance it.

Finally, Table 2 shows that, just like in developed countries, congestion when accessing large cities and between large cities is increasingly common throughout the developing world[5]. This is particularly well recognised in Latin America and Asia, where large projects are designed to promote suburban trains and high-speed intercity trains.

3. FORMS OF PRIVATE SECTOR PARTICIPATION IN TRANSPORT

While there are many ways of classifying PPP contractual arrangements, most international databases keeping track of these contracts tend to classify them in four main categories:

- *Divestitures or asset sale* contracts transfer ownership of the business to the private sector and all risks are hence "privatised"; it can take many forms -- public offerings of shares, or private trade sales of assets themselves;

- *Greenfield Project* contracts are for specific projects such as a new toll road or a port or airport terminal; they cover brand-new investment projects which are commissioned to the private sector [Build-Operate and Transfer or Own (BOT or BOO)] and are among the most common contractual forms. The associated commercial risks tend to be the responsibility of the private constructor while other risks, such as exchange or political risks, can be shared to varying degrees with the public sector through various types of legal instruments, such as guarantees or explicit subsidies. The allocation of these risks in a ring-faced part of a sector is one of the major project finance techniques which have contributed to the development of this form of contract;

- *Service contracts for Operation and Maintenance (O&M) with some investment obligations* are essentially contracts to allow a private operator to manage (i.e. operate and maintain) the service, but do not include investment obligations. These contracts are typically of short to medium duration (2-5 years) and generally the government continues to take on all risks involved in the project except for the management risks;

- *Concessions/licences/franchises* are usually long-term contracts of between 10-30 years, which pass on the responsibility for O&M to a private operator and include detailed lists of investment and service obligations. In this case, the government passes on the commercial risks to the private operator. For many governments, it also has the advantage that it does not imply a politically sensitive transfer of ownership of public assets to the private sector; assets are "rented" out.

In developed countries, asset sales (most obvious in Australia and continental northern Europe) and concessions/franchises (in the UK, southern Europe and Canada) have been the main instruments of private participation in transport infrastructure, but not the only ones. Additional complementary initiatives abound. The best known may be the Private Finance Initiative (PFI), launched by the UK in 1997, to deliver new and modernise existing infrastructure to public services. The initiative is interesting because it has been adapted for the transport sector of other developed and developing countries.

Although it had a wider sectoral agenda, the UK's PFI was designed to stimulate the introduction of private-sector ownership in public services and the transport sector is an integral part of this programme[6]. The PFI in transport has so far been used in projects which are capital intensive and have significant ongoing maintenance requirements. With PFI, the public sector defines what is

required to meet public needs and remains the client throughout the life of the contract. The public sector also ensures, by contract, delivery of the outputs it sets. The private sector takes on the responsibility for providing a public service, including maintaining and enhancing or building the necessary infrastructure.

While significant, the role of the PFI should not be overestimated. It will account for just 11 per cent of total investment in 2003-4 (approximately £4.5 billion) and, from this amount, the transport sector will receive £1.3 billion. The largest transport projects financed through the PFI are the modernisation of the London Underground and the M6 Toll Road.

Other European countries are also increasingly relying on PFIs for transport projects. Germany is relying on private investment (construction, operation and maintenance) to extend the federal motorways. A new law (April 2002) provides the legal basis for implementing a distance-related toll for the use of motorways by heavy vehicles, which will be the finance source for the investment done by the private sector. Ireland also created a PFI and will rely on it to fund some toll roads.

Australia, after selling assets and privatising through long-term leases (for instance, airports), took the lead in setting administrative rules to ensure access to essential facilities, estimating productivity gains and setting benchmarking mechanisms to compare the performance of all operators[7].

In developing and transition economies, concessions are the most common form of private sector participation in transport. Table 3 illustrates that concession contracts accounted for 56 per cent of all transport projects between 1990 and 2001. It also shows that while Latin America and East Asia are the most active in promoting concessions, greenfield projects have been quite successful in East Asia over the last eleven years or so. The Middle East has been the least effective (or maybe the least interested) at building a partnership with the private sector in transport, although there was some activity in private ports. South Asia and Africa come next -- including the successful bi-national railway concession between Abidjan and Ouagadougou. Part of the problem is that in these regions most types of risk levels, not only political and regulatory but also commercial, are high. The ability to pay for transport services in many countries of this region is very modest, implying that tariffs for most passenger services, for instance, have to be quite low. This, in turn, implies rather long-run commitments in order to recover investments.

Table 3. **Types of private sector involvement across countries in transport in developing and transition economies (Number of projects per contract type between 1990 and 2001)**

	Africa	East Asia	Eastern Europe	Latin America	Middle East	South Asia	Total
Divestiture	3	29	17	7	0	0	**56**
Greenfield Projects	11	82	12	60	8	23	**196**
O&M Projects	12	7	5	14	3	2	**43**
Concession Contracts	11	111	9	229	6	16	**382**
TOTAL	**37**	**229**	**43**	**310**	**17**	**41**	**677**

Source: World Bank PPI database.

The distribution of contract types across regions reflects differences in politics, history and risks. Latin America has tended to prefer concession contracts over the other forms because it was believed to be the most effective way of shrinking the fiscal cost of the sector while maintaining the ownership of the assets in the public sector. East Asia, on the other hand, has benefited from the willingness of large local regional financing sources to contribute to expansion projects with high potential payoffs. For the poorest regions, such as Africa and South Asia, greenfield projects were effective solutions to isolate risks to specific activities for which these risks would be manageable. New port terminals, for instance, have often been co-financed by shipping companies. The relatively small proportion of O&M projects reflects the fact that these types of project are usually too small to be picked up by this type of database but, most importantly, the fact that the PPP wave of the 1990s was mainly designed to get the private sector to take on the investment responsibilities. O&M contracts come short on that ground and were hence less attractive at the time.

The evolution of contract types does, however, give a sense that there may be a shift in this view of the world. Indeed, since the Asia crisis, but in particular since the Argentinean crisis, very few new concession contracts have been signed, while the number of greenfield contracts tripled (from 67 to 196) between 1997 and 2001 -- the few concession contracts which were eventually signed had been in the pipeline since well before the crisis. The ongoing dialogue with the major sponsors of these contracts is quite clear: they are no longer willing to take on as much risk as in the 1990s.

This increased risk aversion is quite clearly revealed already in the evolution of the financial structure of listed private projects. According to Correia *et al.* (2004), the leverage rate has been increasing significantly since 1998 for all infrastructure projects in developing countries, but in particular in transport. The debt-to-equity ratio for most transport projects, based on the market value of the assets, is now around 60 per cent, an increase of about 50 per cent over the mid-1990s values. The cost of capital has also been increasing quite significantly. Estache and Pinglo (2004) show that, based on a book valuation of the financing structure (which has tended to be much more optimistic in recent years with the depressed markets), the cost of capital in rail and ports is around 12-13 per cent and the cost of equity around 15-16 per cent on average. Note that for the lowest income countries the cost of equity is significantly higher and reaches close to 20 per cent, at a time when interest rates are at some of the lowest historical levels.

At these costs of capital levels, the willingness to sign concession contracts is bound to decrease and it would seem rational to expect that the demand for contracts requiring no investment commitments, or only investment commitments paced with the ability of the operator to generate cash, is going to surge. At best, even in regions in which concessions tended to dominate, the need to better manage risk is likely to increase the relative share of greenfield projects. This is already the case for Latin America, for instance. Overall, the main beneficiaries are going to be developed countries. With an excess supply of cash on the international financial markets, the willingness of large construction firms to move back to the European market, as the liberalisation speeds up and proceeds well, is already quite obvious for anyone working on project finance.

The evolution of the financing structure, of the cost of capital and of the forms of contracts is not the only source of evidence to assess the evolution of PPP in the sector. A forthcoming study conducted by Guasch (2004), of about 1 000 concession contracts signed during the 1990s in Latin America, suggests that problems have been in the making for a while now. Defining renegotiation as an event in which a concession contract undergoes a significant amendment in any of the following areas -- tariffs, investment plans and levels, exclusivity rights, guarantees, lump-sum payments or annual fees, coverage targets, service standards and concession periods -- Guasch finds that transport concession renegotiations are almost a norm rather than an exception. Indeed, Table 4 shows that 55 per cent of all transport concessions implemented between 1985 and 2000 in Latin America and the

Caribbean region were renegotiated[8]. This percentage implies that more than one out of every two transport concession contracts has been renegotiated[9]. It took as little as 3.1 years before the average transport contract underwent a renegotiation.

Table 4. **Level and speed of concession renegotiation requests in Latin America**

	All infrastructure sectors	**Transportation**
Percentage of renegotiated contracts	30.0	54.7
Average time to renegotiate since award (years)	2.2	3.1

Source: Guasch (2004).

As pointed out by Guasch, not all renegotiations were bad. A renegotiation can indeed be a welfare-enhancing instrument when used to address the inherently incomplete nature of concession contracts. Although some renegotiations were desirable, the high percentage of renegotiated contracts indicates the design, implementation and enforcement of concession contracts will require a careful fine-tuning. The main problems identified by Guasch (2004) include the incompleteness of the contracts, the lack of competitive bidding processes with clear and transparent award processes and the absence of an independent and technically well-endowed regulator. It is interesting to note that in the transport sector, in 57 per cent of cases, the renegotiation was requested by the operator (*vs.* 61 per cent for all infrastructure contracts) and that the government was the sole initiator in 27 per cent of cases, the highest rate across sectors[10].

4. HAVE TRANSPORT DE-REGULATION AND RESIDUAL REGULATION BEEN EFFECTIVE?

There are three main dimensions by which the effectiveness of the reforms can be assessed:

1) efficiency;
2) fiscal;
3) the user viewpoint.

To our knowledge, there is no encompassing quantitative study of the effects of reform in developed or developing countries according to these three dimensions. Only partial evidence tends to be available. This chapter reviews this partial evidence (or the lack thereof) and provides a rough assessment of the overall effectiveness of the reforms based on that evidence.

4.1. Efficiency

While deregulation in the sector allows for significant intermodal competition, which should take care of ensuring efficiency gains to be passed on to the users, natural limits to competition continue to be a major issue in the specific context of transport infrastructures such as port facilities, and for captive shippers who are the main clients of freight railways in many developing countries. A major indicator of the effectiveness of deregulation and of the residual regulation by the public sector should then be the evolution of the efficiency levels of the sector[11]. The correct (or least incorrect!) measure of efficiency comes from the estimation of a production or cost frontier which gives, respectively, a sense of the maximum production possible from an existing stock of inputs and technology, and a sense of the minimum cost possible with the same characteristics[12]. The potential efficiency gains for any operator can then be assessed from the position (and its change in time) of each operator relative to the frontier[13]. The use of efficiency measurements generated from production frontiers is a suitable instrument not only to assess the effectiveness of reforms but also to foster *ex post* competition in transport markets. Indeed, it could induce the implementation of yardstick competition schemes, aimed at giving stronger incentives to the poorest performers to improve their efficiency.

For developing countries, these efficiency gains should be expected to be particularly important, since the initial conditions had been known to deteriorate so dramatically as a result of long-lasting fiscal crises that they facilitated the political support for the reforms of the 1990s. Governments have a variety of objectives when they decide to privatise transport services but the two most important are the need to reduce their fiscal burden and to bring new investment to these services. The quantity and quality of transport services were expected to increase as a result of privatisation, primarily as a result of new investments (greenfield projects or maintenance and rehabilitation of existing infrastructure), but also because the private sector was expected to bring in significant new expertise and fewer political constraints, which should contribute to improving productive efficiency. From a very pragmatic viewpoint, the potential or realised efficiency gains are a good approximation of the extent to which prices could drop in the sector without threatening its financial viability. For passenger services, where tariffs continue to be controlled in many countries of the world, they can be used to assess the maximum subsidy the government should be willing to pay.

For developed countries, the decision to deregulate and restructure may have been less of a pragmatic solution to pressing problems and more of an ideological choice, supported by the majority of the population through elections. They are nonetheless expected to be important as well, in particular in countries where long traditions of political interference with the management of the sector, including input and output prices, tended to distort investment and production decisions.

Despite its importance for the monitoring of the effectiveness of reforms in their sectors, few regulators in developing as well as in developed countries (with the notable exception of the UK and Australian regulators) have taken on the challenge. There are, however, a number of studies to draw to acquire a sense of the potential efficiency gains and the impact of reforms.

The recent relevant literature is reviewed in Coelli *et al.* (2003) and more detailed surveys on ports and railways, including an evaluation of the literature on European case studies, are available from Gonzalez and Trujillo (2003). Overall, efficiency has tended to improve on average, according to most studies, although the difference across operators continues to be sometimes very significant. Most of this literature, however, is not centred on the impact of a reform on the efficiency levels. It tends to focus instead on comparative assessments, which do not account as well as expected for differences in institutions, ownership, regulatory regimes or overall political and economic conditions. One of the few exceptions is a recent paper (Friebel *et al.,* 2003) which estimates the effects on efficiency that the European Commission's Directive[14] imposed on the railways industry. Using a

World Bank database[15], this paper argues that reforms -- separation of infrastructure and regulation, third-party access and independent regulation -- increased efficiency. However, a striking result is that multiple reforms introduced simultaneously have neutral effects, but sequential reforms improve efficiency. For developing countries, the estimation of efficiency and production frontiers is seriously limited by data restrictions. Two exceptions are studies of Latin American reforms. Estache *et al*. (2002b) shows that the Argentinian and Brazilian railways generated significant efficiency gains. Estache *et al.* (2002a), estimated average annual efficiency gains in the Mexican port sector, since the reforms were implemented, of around 3 per cent. This paper highlights the fact that efficiency estimation is feasible and transport regulators have much to win from it; especially because it allows the promotion of yardstick competition, which is crucial to ensure that whatever short-term gains are achieved from reform can be sustained and reinforced over the long run.

In sum, the evidence suggests that the efficiency gains have been significant. Their specific source, however, is less documented than one would have expected.

4.2. Fiscal

From a fiscal viewpoint, whenever a reform includes a privatisation component, in general the fiscal impact has been positive, simply as a result of payment from the sale in case of divestiture or from the rental of infrastructure in the case of some concessions. This has been fairly well documented during the 1990s, in particular in the context of the eastern European reforms. But most of these studies fail to indicate a few crucial points. First, the fiscal gains achieved from sales or rental tend be short-term gains. A recent study on the Latin American experience, which combines debt reductions, sales and rental fee gains, shows that, while the reforms were indeed effective in maintaining low levels of public investment in the sector, the demand on recurrent expenditures increased through renegotiations [Campos *et al.* (2003b)]. Most of this increase has come from the demand for subsidies for passenger transport. Indeed, as politicians are reluctant to increase tariffs when costs increase, the only way for private operators to maintain financial viability for a given investment level is to benefit from operating subsidies.

Second, the reduction in total public expenditures is not consistent with the needs and simply implies increased rationing, because (i) the private sector has not picked up as large a share of the demand as was expected and, hence, (ii) the public sector overshot the reduction in its commitment to the sector. This was not only because the expectations from the private sector were too high but also because the infrastructure in general was one of the public expenditure categories targeted for cuts under fiscal adjustment programmes, without accounting for the needs and the other macroeconomic implications. According to Calderon, Easterly and Serven (2003), the infrastructure sector (including transport) contributed on average about 50 per cent of the total fiscal adjustment in the region. This translates into a 1 per cent lower long-term growth rate for the region; and its own fiscal consequences, since lower growth means lower tax revenue.

In sum, the short-term fiscal gains have generally been positive but "the jury is still out" on the long-term gains. The evidence indicates that fiscal losses are likely to emerge from the mismanagement of the reform from: (1) a return of operational subsidies; (2) the negative fiscal effects associated with lower growth, due to overshooting the reduction in public investment in the sector.

4.3. The user viewpoint

From the viewpoint of the users, two dimensions need to be considered: price and quality[16]. To address these dimensions, it is useful to distinguish between passenger and freight transport because the concerns are, to some extent, different.

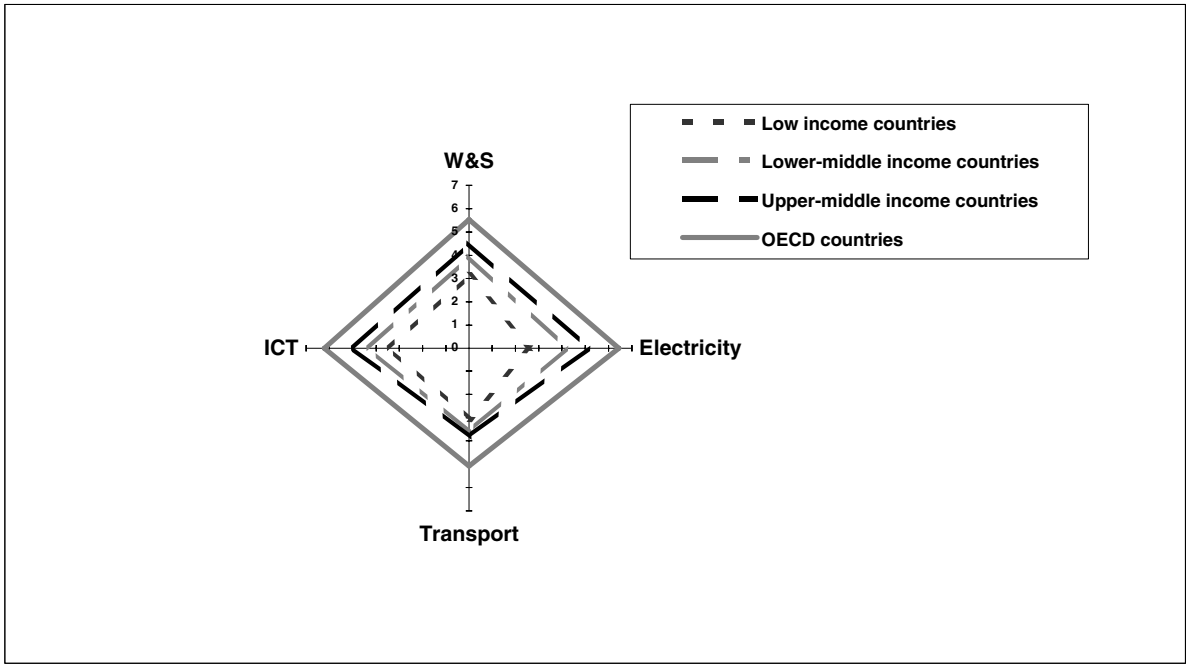
Starting with passenger transport, the main infrastructure issue emerging from the 1990s may be the financing of intercity and suburban railways, where private operators seem to have problems in maintaining some degree of financial viability without government support on many lines. Indeed, the control of fares is a regular theme in the media around the world and in the poorest countries is a major source of conflict between user representatives, operators and the government. Similarly, in many countries, conflicts have tended to emerge from the closure of unprofitable rural lines, leaving some users at least worse off and partly offsetting the fiscal savings derived from reduced service obligations and the gains from other quality improvements (reliability, security, etc.) which have resulted from many of the reforms.

Urban transport is less of a problem because it is largely competitive (buses, trains, taxis and similar services all compete for the same passengers in many countries). With the exception of subways, where financing is also needed, there is little the public sector needs to do except maybe to regulate quality, ensure that competition takes care of prices and facilitate intermodal integration. The main issue may indeed be the need to better manage traffic to address safety and environmental concerns. This is largely being done, slowly but surely, as the concern for pollution has become a worldwide policy item and as, at least in the developing world, the concern with the lack of safety, associated with unmanaged competition resulting from excessive liberalisations, is now leading governments to consider hybrid solutions.

From the viewpoint of freight transport, the story emerging from the 1990s' reforms is somewhat more complex. Most infrastructure privatisations have followed a price-cap type of regulatory regime, at least in developing countries. Because these caps have seldom been binding and the competition with the trucking industry has been quite strong, it is unlikely that the average users may have been penalised by the reform. The main residual concern may come from captive shippers (this is discussed later in the paper). As for quality, there is mixed evidence. Visible quality has generally improved (punctuality, safety, etc.). Less visible (environment) quality has improved but not as much as expected. Highly visible accidents in the UK have provided highly mediatised (but hardly scientific) evidence of the problem and similar stories are emerging in developing countries. A recent study of the use of under-investment in safety by Brazilian railway operators confirms this [Estache *et al.* (2004c)].

Overall, the bottom line may be provided by a perception index collected by the World Competitiveness Report. This index gives a sense of what commercial users of transport services feel about the overall quality of infrastructure services. Figure 1 reports the data for each sector and a grouping of countries which allows a double comparison. First, transport is clearly the worst rated of all sectors. Second, this is true across country groups. This suggests that the reforms in the transport sector have in general not been able to generate the level of satisfaction observed for other sectors.

Figure 1. **Commercial perception of transport quality as compared to other infrastructures (per country groups)**



Note: Index based on quality perception: 1=underdeveloped; 7=as developed as the world's best.

4.4. Summing up

The main message which flows from this very brief overview is a relatively simple one. Deregulation and restructuring did, in general, improve efficiency and did improve prices on average, except maybe for captive shippers, but it came at a fiscal cost and possibly at a long-term quality cost — which may include the lack of a multimodal vision in the design of the sector-specific reforms. This fiscal cost is a problem which should have been addressed from the start of the process and which was probably not addressed because most of the fiscal objectives of the reforms were set by teams insufficiently familiar or concerned with the specifics of the sector. The quality and related captive issues are evidence that the residual regulatory role may not yet have been addressed sufficiently seriously.

5. TOWARDS A RE-REGULATION OF THE SECTOR?

While it is clear that competition should be the dominant policy aimed at in this sector, there is enough evidence to point to the need for an improvement in the residual regulation to be provided by the public sector in many countries. Based on an informal (and largely personal) observation of the main issues emerging from the 1990s experience, this paper focuses on:

- Prices: including average prices and tariff structures for monopolies;
- Quality visible (timetables, safety…) and less visible (environment…);
- Access rules for captive shippers and interconnected networks;
- Rebundling of a sector and hence decrease in intrasectoral competition;
- Additional stimulus through yardstick competition;
- The stimulus to multimodalism;
- Local *vs*. global competition issues.

5.1. Price regulation

There are two main types of pricing issues emerging for regulators in the context of the residual regulation of transport infrastructure. The first is the need to recognise that PPP changes the pricing rules in a dramatic way. Indeed, for private investors to consider partnering with the public sector, in a sector regulated under price caps or some type of incentive-based regulatory regime, the business needs to generate a return at least as high as the cost of capital they are facing. The estimation of the cost of capital is usually done at the time the project finance of the reform is being worked on by the privatisation team, and should be revised on a regular basis by the sector regulator. This is, however, only done by the best regulators in the world (generally, the Australian and the British) in preparation for scheduled tariff revisions. The only (to our knowledge) transport regulator who has so far gone through this kind of process in the developing countries is the Mexican port regulator. This is quite unfortunate because investors in developing countries are much more sensitive to changes in risk levels, including political and foreign exchange risks, and hence expect their tariff to reflect those changes as systematically as possible. One of the major adjustments needed to the PPP model implemented during the 1990s is getting regulators to organise themselves to be able to estimate average tariffs in a way which gives an incentive to operators to commit for the long run. Regulation matters, and hence regulatory institutions matter too; yet the development of the regulatory capacity of the sector continues to lag the development of the equivalent capacity in other public services.

The second pricing issue, which has tended to be underestimated, has to do with the tariff structure. Indeed, one of the often underestimated dimensions of the interaction between restructuring and fiscal concerns with incidence on prices is the existence of cross-subsidies which can be built into the restructuring design. In the airport sector, for instance, it is usually the case that a few airports concentrate most of the traffic, domestic and international, while other airports have volumes of operation which make them financially unsustainable. Under the operation of a public firm, implicit cross-subsidies are the norm to keep airports open which are considered necessary or strategic. Reformers may opt to sell or concession all airports in a package, instead of concessioning each of them separately and designing an explicit subsidy scheme (in 1998, Argentina concessioned 32 airports to only one private consortium). By keeping the implicit cross-subsidies in place, the government does not need to make explicit subsidies to private operators, which may help implement the reforms but may lose the possibility of increasing competition between markets. Moreover, by not tracking the level of cross-subsidies explicitly and transparently, the regulators may leave an extraordinary rent to the monopolistic operator of the sector.

Overall, this means that serious regulatory accounting guidelines are needed, as well as consultation processes to discuss and implement asset valuations, assessments of potential efficiency gains and economic and financial models to be used by regulators in their transparent decisionmaking.

5.2. Quality regulation

In designing and developing privatisation packages for transport services, it is essential to improve the accounting for quality alongside the traditional quantity and price aspects. Quality includes service level dimensions, such as journey time, service frequency or reliability, as well as non-service level dimensions such as safety and environment.

The rationale to consider quality as an important dimension in the redesign of regulation rests on the evidence of its "adjustment lever" nature. Private concessionaires may take advantage and reduce the quality of the service provided because they may find it commercially convenient to save on costs or because informational advantages and the difficulty of monitoring service levels make it easy not to comply with pre-agreed quality-related standards. Lowering quality is often the way private operators "adjust" to increased competition from other operators, or the way an operator working within a monopolistic framework (e.g. under a licence agreement as the sole operator in a given route) may exploit monopoly power and adjust service quality according to demand responsiveness. Thus, quality matters both in competitive and monopolistic structures where transport services are being privately provided.

The presence of private operators in partnership with the public sector will require a reallocation of responsibilities regarding safety and the environment. Many safety and environment standards tend to be set by national legislation and frequently fall under the responsibility of national bodies, such as a health and safety executive, a ministry of environment or an environment agency. A particularly important aspect of setting safety standards is provision for third-party liability insurance, in particular, the need to assess the amount of cover necessary, the extent to which operators are able to insure at reasonable terms in the insurance market and the amount of self-insurance allowed. Dispute resolution in the case of a serious accident occurrence among the parties involved, regarding liability and cost recovery allocation, should also be carefully assessed. Generally, environment standards will be set and monitored by environmental authorities based on standards set nationally or supranationally. Nevertheless, the concession/franchise awarding authority may use environmental quality indicators in the valuation of bids; and the regulator may be delegated to monitor compliance with environmental standards on behalf of the environment agency. In those circumstances it will be a condition of the operator's licence to show commitment by specifying a plan and the management means to achieve it.

5.3. Access regulation and competition policy

In the provision of transport services, some facilities are necessary inputs in the production process. Such is the case of airports in air transport, loading docks in maritime services and rail tracks in railway services. When antitrust authorities include only one facility in a relevant geographic market[17], this facility is considered essential. In practical terms, an essential facility implies that it is not economically reasonable to duplicate the facility. If only one facility must be used by all service providers, the establishment of rules to determine how, when and at what prices access is allowed is the key to guaranteeing effective competition in the market.

Unless access pricing rules are defined before the business is passed on to private operators, it is clear that rents are being created which are harmful to users. This has long been an issue in the US, it is an issue in the UK and it has proved to be an issue in most developing countries where the need to transfer the business to private operators is often so pressing that there is little time to work out the demanding details of access pricing.

In general, when reformers identify an essential facility, they create a regulator which sets access tariffs and other conditions designed to avoid undue discrimination. The relevant question is whether the regulator is strong enough to deter the monopolist from abusing its dominant position. Of particular relevance are the cases where the monopolist can compete with the providers of services in those segments that are deregulated. In the case of railways, can the operator of the rail tracks own and compete with other companies which own trains and operate on those tracks?; can an airport operator own an airline which uses that airport as a hub and compete in a deregulated market?

Without access price regulation, it is clear that the operator of the essential facility can extract all the rent from the operators which use the facility. However, when there is access price regulation but the concession contract or the privatisation law says nothing about the possibility of vertical integration, the operator of the essential facility has strong incentives to control an operator downstream, and implement a rising rivals' cost strategy to exclude all competitors and become a monopolist both upstream and downstream. Even when prices are regulated, there is a wide variety of conduct which can be implemented by the operator of the essential facility, whose purpose is to reduce competitors' quality of service -- which is equivalent to increasing their costs.

When restructuring a transport sector, what can governments do to establish an adequate access regime? One extreme solution is a strict vertical separation between control of infrastructure and users of the infrastructure. The disadvantage of this alternative is the likely loss of economies of scope and adequate information to plan infrastructure investments[18]. If a government decides not to prohibit vertical integration, to avoid future attempts to vertically integrate through access discrimination, it should implement an adequate open access regime, complemented with rules of accounting separation, and it should set up a regulator with enough resources to be a credible watchdog of competition. Some countries, New Zealand being the best example, prefer not to impose any restriction on the market structure and rely on the competition agency to avoid the application of any anticompetitive conduct.

The airport sector provides an example of the variety of approaches adopted with respect to the degree of vertical integration. Australia and Chile have explicit rules against vertical integration: Australia limits an airport operator's ownership stake in an airline to 5 per cent and, in Chile, the bidding guidelines for airport concessions specify that the infrastructure concessionaire cannot have decisive influence over the administration or management of companies offering air transport services. Although the private sector does not have a significant role in the airport sector in continental Europe, the European Commission also recognised the potential problems of vertical integration in its analysis of the proposed merger between Air France and Sabena in 1992[19].

International experience -- including the recent ruling by the Argentine antitrust agency, which rejected a merger attempt by the operator of 32 airports to buy an airline -- suggests that developing countries restructuring transport markets should include an explicit prohibition on vertical integration (impose vertical separation) between the owner or operator of the essential facility and its users. This regulatory approach has several major advantages: it keeps the costs of monitoring and information gathering low, eliminates the incentives to transfer market power to the competitive segment, reduces conflicts between the regulatory and competition agency and provides certainty to transport service providers in the competitive segment.

5.4. Rebundling

In practice, restructuring generally implies some degree of unbundling of the activities performed in each subsector and it consists in an actual disintegration of the monopoly into various business units. The restructuring can be horizontal, creating companies which deliver very similar activities. This is common in railways (many companies providing freight or passenger services), ports (companies operating different terminals within one port or one company operating each of many ports in a given region) and airports (one airport per private operator in a given country). Unbundling can also be vertical and determines the extent to which a single firm can participate in different vertically related stages of production. For instance, in the airport sector, vertical unbundling implies that the airport operator cannot hold a controlling stake in an airline which operates from that airport.

Despite the initial desire to cast the market structure to increase competition in the market, operators seem to be increasingly tempted to subsequently try to concentrate horizontally or vertically. An example of unbundling followed by a process of rebundling is given by the port sector in Argentina[20]. In the early 1990s, Argentina restructured and privatised its port system. As part of this process, Puerto Nuevo's six terminals (the largest port in Argentina) were offered to the private sector under long-term concessions. The Government showed a concern for competition throughout the concession process, imposing conditions which would result in a market structure capable of sustaining competition: bidders were allowed to bid for more than one terminal, but they had to express a preference and could be awarded only one. The conditions imposed by the Government implied a prohibition on horizontal mergers between terminals. However, reformers were not concerned about the vertical structure of the market. Thus, P&O, a multinational firm which operates both terminals and a maritime shipping company, was awarded two terminals and, in 2001, the Argentine Antitrust Commission approved the acquisition of a terminal by Maersk Sea Land.

Only six years from the start of the concession, in 2000, probably because of excess capacity, three terminals proposed to merge. Allowing this merger would imply a change to Puerto Nuevo's regulatory framework which explicitly prohibited mergers between terminals. When should governments allow a rebundling process to take place? Is it advisable to allow mergers between port terminals? Providing an answer to these questions involves debating about the convenience to commit to the text of a regulatory framework or to choose a flexible scheme to adapt to a changing economic environment. Of particular importance is the recognition that the cost of capital has increased significantly and that rebundling may be a rational strategy to maintain it at a competitive level. In Argentina, the Government opted to change the regulatory framework and, subject to the ruling of the antitrust agency, allow horizontal mergers in the port sector. In the port sector, as in any other transport market, antitrust agencies must perform a cost-benefit analysis to assess the effect of mergers on competition and consumer welfare.

More generally, in transport markets, firms will cite efficiency gains as the main reason to merge, achieved through the potential for economies of scale, the possibility of rationalising resources and the presence of actual or potential intra- and intermodal competition. On the other hand, antitrust agencies must carefully consider the potential harmful effects of mergers on competition. These are the consequence of barriers to entry, strategic (an example applied to the port sector is signing long-term contracts with all shipping companies to prevent the construction of a new terminal or nearby port) or legal.

5.5. Additional stimulus through yardstick competition

The deregulation of the sector has often resulted in the creation of local or regional monopolies in airports, rail and ports. This overall reform process in the sector is generating a significant increase in the number of combinations of public and private operators involved in these businesses. This new market structure is providing a unique opportunity to introduce yardstick competition in transport markets in such a way that it could become an effective regulatory tool in this sector, as it has become in the utilities sector. Indeed, the formal introduction of interregional competition in these sectors could be designed to promote efficient management of services which are otherwise operating under local monopolies. Yardstick competition allows the performance of the various operators of a given sector to be compared over time, continuing the competitive pressure. In fact, while no country yet has formal experience with this tool in the transport sector (with the exception of the Mexican port sector), many new regulators recognise its importance to ensure sustainability and to improve the initial gains achieved through the restructuring and privatisation process.

This will, of course, require a much stronger commitment to regulatory processes and procedures, which allows the generation of consistent international databases, as has been the case in water or energy for instance.

5.6. The stimulus to multimodalism

One of the chief failures in the design of the 1990s' reform in developing countries, and a major gap with respect to achievements in developed countries, is that reforms followed a piecemeal approach without an integrated vision for the sector. Every subsector had its PPP team and seldom talked to each other, while ignoring the possibility of favouring intermodal integration. Yet, high logistics efficiency tends to require competitive incentives for service providers to continuously innovate and seek out lower cost combinations of transport services in a co-ordinated way, in order to minimise confusion and increase the predictability of the environment in which transport operators make investment decisions.

For many developing countries still lagging, it might be preferable to simply adopt the relevant standard international statute(s) to ensure international compatibility. To do so, governments need to clarify the public service obligations of core transport service providers -- such as port, rail and intracoastal and water barge operators -- including the obligation to provide open access to their service networks to third-party users, such as commercialisation agents (e.g. freight forwarders, multimodal transportation operators). These agents are qualified to integrate the core transport modes into door-to-door service packages under a single bill of lading. Several types of rules and standards, promoting "connectivity" among stand-alone services and "interchangeability" of intermodal equipment among carriers, will be required to ensure that an integrated network will work effectively, and to give multimodal service providers a much better basis for understanding how they can develop markets for multimodal services.

5.7. Local *vs.* global competition issues

In spite of all the efforts made by reforming governments, it seems that competition for the market is not working that well, in most developing countries at least. Typically, the number of serious bidders for a concession or a greenfield project in transport infrastructure is not much higher than two or three — there are, of course, examples with more players but a large number of players is not the norm in infrastructure concessions. Technically, this is competition, of course, but it is not

impressively strong competition. The fact is that if it were not for the foreign bidders, it would often be the case that the sector would remain in the control of a very limited number of local players. Most typically, these players are local construction companies which have historically been winning all the procurement contracts offered by the public monopolies to implement their investment programmes. The opening of the sector is at least ensuring that, even if specialised foreign operators are not interested, foreign construction companies will be able to compete with the local construction firms in putting together bids for these contracts.

The simple fact is that the transport sector is a highly concentrated sector, internationally as well. This issue has been raised for quite a while now in maritime transport, where shipping companies are increasingly integrating vertically to take over the container terminal business. There is indeed casual evidence of increases in dedicated terminals, in which one shipping company monopolises at least part of the port infrastructure. This is the case in Antwerp, with MSC, and in Rotterdam with Maersk, but also in Santo Domingo and in Buenos Aires. Sometimes there are good reasons to reduce the number of terminals, including the increased cost of capital in emerging economies, which make mergers an effective risk-management instrument, as mentioned earlier. But with the ongoing wave of mergers or alliances between maritime companies, there is reason to wonder about the risk of effective inter-port or inter-terminal competition which results from the reduction in the number of suppliers of maritime services. Similar issues arise in the airport management business, where the same 5-10 usual actors seem to be present at all bids[21].

The least recognised indicator of concentration at the global level may be in the construction sector, which is at the core of all new infrastructure projects. Working with statistics published by Public Works Financing, the construction deals on all of infrastructure (transport + utilities) are highly concentrated in very few hands. Indeed, six companies share 50 per cent of the market and 16 companies share 90 per cent of the market. Spanish companies (Dragados, Ferrovial, Abertis, OHL, FCC, Acciona and Sacyr), accounted for 52 per cent of all new concessions and PPP projects of over US$50 million, under construction and signed between 1985 and 2003. British companies (John Laing, AMEC, Balfour Beatty and Alfred McALpine) accounted for 14 per cent as well as French companies (Vinci-Cofiroute, EGIS, Bouygues, Alstom), while Australia accounted for 9 per cent (Macquarie). In sum, ten companies share about 90 per cent of the market! The concentration is actually quite impressive, since many of these companies have a certain degree of regional specialisation.

The point is that effective *ex ante* competition tends to be modest, even in some of the best organised auctions, and that the opening of trade in services is not changing the picture much. But there is more to be squeezed out of the experience. In particular, it is interesting to look into the specific nature of the bidding companies. In many cases, the bids are made by consortia combining local and foreign companies. There are, of course, gains in terms of transfer of know-how and reductions in the overall risks associated with these strategies. But the gains may sometimes have to be weighted against the cost[22] induced by lesser competition.

6. CONCLUSIONS

Three main points emerge from the paper. The first is that deregulation has generally had different effects in developed and developing economies. Project sizes are larger for developed countries. The relative contribution to needs is larger as well (in most developed countries, installed capacity is usually much larger already and marginal projects have tended to have a much higher marginal impact). The fiscal payoffs have generally been good in the short run but less predictable in the long run, certainly in the case of developing countries, often as a result of contract renegotiations. As for the users, the payoff from reform is somewhat different for passenger and freight transport. Passengers have seen improvements in the quality of services, have more options to choose from and often face relatively low prices, but often because the government has kept a tight control — with important fiscal consequences. Freight shippers are also better off overall as a result of deregulation and restructuring. The potential for improvement, however, is still strong. Few countries have reached their potential in terms of modal integration. Moreover, safety in some cases continues to be an issue and the increased concern for security resulting from September 11 is unlikely to smooth things out.

The second point is that the odds of the private sector playing a strong role in the transport infrastructure of all countries of the world are low. The concentration rate of projects is significant. The OECD and a handful of developing countries is where the private sector has gone and will continue to go, at least for a while and until developing countries manage to generate politically viable projects (from the user viewpoint), with returns sufficiently high to match the high cost of capital linked to the perception of high risk associated with many developing countries. Many of the "not chosen" countries have made strategic mistakes in cutting public sector commitments, in the hope that the private sector would do it, and are now forced to "play catch-up" for supply to meet demand.

The third, and maybe the most important point is that the policy agenda associated with the follow-up to the strong commitment to deregulation and increased private sector participation, is far from being small. The issues reviewed in Chapter 5 are often complex to address. The corrections needed to the reform path require a strong political commitment at the national and international levels. This commitment is needed because addressing these issues implies strong redistribution of rents, which politicians are seldom willing to implement. Ultimately, the sustainability of the reforms and the ability of the public sector to use money more effectively in leveraging private money (including guarantees and subsidies) will depend significantly on this political commitment to effective regulatory and competition[23] policy. History suggests that fine-tunings are often more difficult to implement than large reforms. If transport ministers are to endorse this emerging policy agenda, a new hybrid model of PPP will emerge, with a significantly larger positive impact for users, operators and for current as well as future taxpayers!

NOTES

1. A similar story is told for the evolution of public-private partnerships in the bus sector, with illustrations for Chile and Colombia in Estache and Gomez-Lobo (2004).
2. For an overview of the core issues in restructuring, privatisation and regulation of the transport sector, see, for instance, Estache and de Rus (2000).
3. 47 per cent of projects and 51 per cent of total funds.
4. This kind of back-of-the-envelope estimation shows why the share of financing transport infrastructure by the private sector is likely to be much lower than the 20 per cent estimated by DFIF (2002) for the whole of the infrastructure sector.
5. Highways dominated private activity in toll roads, accounting for 94 per cent of the investment in 1990-2001. Most of the toll road projects connected major metropolitan areas, where large traffic flows supported the projects' financial viability.
6. According to HM Treasury (2003), up to 1997, the transport sector had suffered from consistent neglect and a highly damaging lack of investment in infrastructure.
7. Studies specific to airports, ports and railways can be obtained from the Australian Competition and Consumer Commission (www.accc.gov.au)
8. The empirical findings in Guasch (2004) rely on a dataset of more than 1 000 concessions granted in the Latin American and the Caribbean regions during 1985-2000.
9. In relative terms, the number of contracts renegotiated are much higher than in the electricity sector and the average of all sectors, but it is lower than the percentage in water and sanitation, see Guasch (2004).
10. A fifth one, not addressed here, is the macroeconomic impact of the reforms specific to the transport sector. While there is a growing volume of publications on the positive effects of the sector on growth levels or growth convergence, there is hardly any study of the marginal effects of transport reforms on these variables. A recent approximation comes from the *ex-post* evaluations of social rates of return done of World Bank projects focusing on policy reforms in the sectors. These rates of return turn around 35-40 per cent (see Estache (2004).
11. For a useful and early overview of the US experience, see Winston (1993).
12. Most transport ministries or regulatory agencies tend to focus on partial performance indicators, relating one input to one or more outputs (labour or capital productivity). These are unfortunately often misleading and quite useless for quantitative regulatory decisions.
13. For more details, see Coelli *et al.* (2003).
14. EC Directive 91/440.
15. Railway Performance Database, The World Bank Transportation, Water and Urban Development Unit. http://www.worldbank.org/transport/rail/rdb.htm. Countries covered are: Austria, Belgium,

Denmark, Finland, France, Germany, Italy, Netherlands, Portugal, Spain, Sweden and the United Kingdom.

16. The importance of quality in the context of transport markets is emphasized by Gomez-Ibanez (2003) in his study of the experience of Sri Lanka.

17. A relevant geographic market is the minimum area where a hypothetical monopolist can profitably increase prices. In other words, it is the minimum area where there is no effective competition because consumers cannot demand services from other suppliers.

18. For instance, in the case of airports, airlines have ample information on trends in air traffic demand which is needed to plan infrastructure investments. But, with an adequate information regime embedded in the regulatory framework, these benefits can be obtained without vertical integration.

19. European Commission, Air Fance-Sabena, IV/M 157.

20. See Serebrisky and Trujillo (2003) for more details.

21. For more details and examples, see Estache (2001).

BIBLIOGRAPHY

Alexander, I., A. Estache and A. Oliveri (2000), A Few Things Transport Regulators Need to Know About Risks, *Utilities Policy*, Vol. 9, No. 1 pp. 1-13.

Benitez, D. and A. Estache (2004), How concentrated is the control of the infrastructure subsectors in the world?, mimeo, The World Bank, Infrastructure Vice Presidency.

Calderon, C., W. Easterly and L. Serven (2003), Latin America's Infrastructure in the Era of Macroeconomic Crises, in: Easterly, W. and L. Serven (2003) (eds**.),** *The Limits to Stabilization -- Infrastructure, Public Deficits and Growth in Latin America*, Stanford University Press, pp 21-94.

Campos, J., A. Estache and L. Trujillo (2003a), Processes and accounting matters for regulators: Learning from Argentina's railway privatization, *Journal of Network Industries.*

Campos, J. A. Estache, N. Martin and L. Trujillo (2003b), Macroeconomic Effects of Private Sector Participation in Infrastructure, in: Easterly, W. and L. Serven (2003) (eds.), *The Limits to Stabilization – Infrastructure, Public Deficits and Growth in Latin America,* Stanford University Press, pp. 139-170.

Correia, L., A. Estache and S. Jarvela (2004), Evolution of the financial structure of infrastructure projects in developing countries, mimeo, The World Bank, Infrastructure Vice Presidency.

Coelli, T., A. Estache, S. Perelman and L. Trujillo (2003), *A Primer on Efficiency Measurement for Utilities and Transport Regulators*, The World Bank Institute Development Studies, Washington, DC.

DFID (2002), Making connections: infrastructure for poverty reduction, mimeo, available at *http://62.189.42.51/DFIDstage/FOI/dc/7mar02_making_connections.pdf*

Drewry Shipping Consultants (2002), Global Container Terminals -- Profits, Performance and Prospects, London.

Estache, A. and M.E. Pinglo (2004a), What is the cost of capital for infrastructure projects in developing countries?, mimeo, The World Bank, Infrastructure Vice Presidency

Estache, A. (2004b), Do ex-post evaluations of the social rate of return on transport projects match the results from econometric studies?, mimeo, The World Bank, Infrastructure Vice Presidency.

Estache, A., S. Perelman and L. Trujillo (2004c), Towards quality adjusted measures of efficiency gains in freight railways: evidence from Brazil, mimeo, The World Bank, Infrastructure Vice Presidency.

Estache, A., M. Gonzalez and L. Trujillo (2002a), Efficiency Gains from Port Reform and the Potential for Yardstick Competition: Lessons from Mexico, *World Development*, Vol. 30, No. 4, pp. 545-560.

Estache, A., M. Gonzalez and L. Trujillo (2002b), What Does Privatization Do for Efficiency? Evidence from Argentina's and Brazil's Railways, *World Development,* Vol. 30, No. 11.

Estache, A. (2001), "Global Competition Issues in Infrastructure Services Trade", The World Bank, mimeo.

Estache, A. and G. de Rus (eds.) (2000), Privatization and Regulation of Transport Infrastructure: Guidelines for Policymakers and Regulators, The World Bank Institute Development Studies, Washington, DC.

Friebel, G., M. Ivaldi and C. Vibes (2003), Railway (De)Regulation: A European Efficiency Comparison, IDEI Report No. 3 on Passenger Rail Transport, Institut d'Economie Industrielle, University of Toulouse.

Gomez-Ibañez (2003), *Regulating Infrastructure: Monopoly, Contracts and Discretion*, Harvard University Publications, Boston.

Gomez-Ibanez, J.A. and J.R. Meyer (1993), *Going Private: The International Experience with Transport Privatization*, The Brookings Institution, Washington, DC.

Gonzalez, M. and L. Trujillo (2003), Medidas de eficiencia en Ferrocarriles y Puertos: Una revision de la literatura, mimeo, Universidad de Las Palmas de Gran Canaria.

Guasch, J.L. (2004), *Granting and Renegotiating Infrastructure Concessions*. *Avoiding the Pitfalls*, The World Bank Institute Development Studies, Washington, DC.

Kennedy, D. (1997), Regulating Access of the Railways Network, *Utilities Policy,* Vol. 6., No. 1, pp. 57-65.

Public Work Financing (1995, 1998, 2003), International Major Projects Survey, *Public Works Financing*,. October.

Serebrisky, T. and L. Trujillo (2003), Market Power: Ports. A Case Study of Postprivatization Mergers, *Viewpoint* 260, The World Bank, Washington, DC.

Serebrisky, T. (2003), Market Power: Airports. Vertical Integration between Airports and Airlines. *Viewpoint,* 259, The World Bank, Washington, DC.

Trujillo, L., E. Quinet and A. Estache (2002), Dealing with Demand Forecasting Games in Transport Privatization, *Transport Policy,* Vol. 9, No. 4, October.

Wilson, T. (1999), The Private Finance Initiative, *Infrastructure Journal*, Winter, pp. 35-43.

Winston, C. (1993), Economic Deregulation: Days of Reckoning for Microeconomists, *Journal of Economic Literature,* 31, pp. 1263-1289.

YARDSTICK COMPETITION FOR TRANSPORT INFRASTRUCTURE SERVICES

Dominique BOUF/
Julien LÉVÊQUE

Laboratoire d'Economie des Transports (LET)
Lyons
France

TRANSPORT SERVICES: THE LIMITS OF (DE)REGULATION - ISBN 92-821-2345-6 - © ECMT, 2006

YARDSTICK COMPETITION FOR TRANSPORT INFRASTRUCTURE SERVICES

SUMMARY

Lyons, February 2004

INTRODUCTION

Infrastructure provision, and notably railway infrastructure provision, is greatly evolving, even if the reform movement is not very fast. In a few decades, the railways have changed markedly in the US, Japan, Africa and South America, and have begun to reform in the EU. The common feature of those reforms is to try to transform the former "protected" monopolies into more competitive or at least more efficient companies. In this general movement, the EU took an original approach, which implies the separation of infrastructure and operation. It follows that infrastructure has to be regulated *per se* and not as a component of an integrated system.

The hypothesis underlying this report is that yardstick competition might be an efficient way to regulate infrastructure provision, at least in its core activity: maintenance. But to support this hypothesis, we need to present yardstick competition quite extensively.

Thus, this report is organised as follows:

- The first chapter gives a brief overview of the regulation problem in order to present yardstick competition in the context of the various ways to regulate firms or utilities;
- The second chapter presents yardstick competition more precisely with a theoretical approach;
- The third chapter provides some examples of the use of yardstick competition in various industries;
- The fourth chapter addresses the main point of the report: is yardstick competition possible and desirable for infrastructure provision?

1. A BRIEF OVERVIEW OF THE REGULATION PROBLEM

In this paper, regulation will be considered as the choice of institutions and rules necessary to achieve the policy objectives which are unattainable by market forces. We present this overview of regulation by first addressing the question of the desirability of regulation, together with the main problems associated with regulation. We will then present the main choices offered to regulators.

1.1. In some cases, regulation is both desirable and problematic

We will first present the main reasons why regulation may be desirable, before looking at the chief difficulties encountered by the regulators.

1.1.1 Why regulation could be desirable

Given the definition of regulation, it follows that regulation is needed where market inefficiencies are found. Thus it is understandable to find market power, opportunism, bounded rationality and externalities among the reasons which make regulation desirable.

1.1.1.1 Market power

With increasing returns, the market forces lead to monopoly. A monopoly may be efficient, as there is no reason why it should waste money, but the optimal prices for a monopoly do not coincide with what would be optimal for society -- which, according to the common wisdom, consists in marginal cost pricing. There has been much discussion and some strong disagreement on the marginal cost pricing rule, but whatever the good price is, there is no reason why it should coincide with the monopoly price, which only maximises the revenue of the monopolistic firm. Thus it is desirable to regulate the prices of a monopoly or to ensure that the monopoly is contestable (this will be discussed later among the options for the regulator).

Regarding transport infrastructure, we are obviously facing some kind of monopoly. But it is noteworthy to add that parallel competition may occur on certain routes. Notably, this is the case for US railways on important routes. This is also the case in countries where tolled motorways coexist with free highways, and thus different routes with diverse characteristics may be offered. Nevertheless, infrastructure is, by itself, a spatial monopoly. By this, we mean that unless an infrastructure is saturated, the economic rationale is to have only one infrastructure between two points. Hence, unless access to the infrastructure is free, there is market power.

1.1.1.2 Opportunism

Opportunism can be defined as the tendency for one economic agent to profit from the occurrence of unanticipated circumstances. More precisely, we can refer to the definition used by Klein *et al.* (1978): once a specialised asset has been set up, there is a tendency by the user of this asset to appropriate part of the quasi rent created by the existence of this specialised asset, to the detriment of the asset's owner. If vertical integration is not possible or desirable, some contracts are necessary between the owner and the user(s) of the asset, but those contracts are by nature incomplete and thus there is a need not only for enforcement but also arbitration and, finally, regulation.

Concerning transport infrastructure, we need to distinguish between road and rail. Tolled motorways can be operated by the State or by franchised companies. But the prices clearly have to be regulated, because of what has been said concerning the market power inherent with infrastructure. For railways, in the case of vertical separation (between infrastructure and operation), any kind of franchising gives ample room for opportunism. In fact, the contract with the operating company cannot be complete and thus many unforeseen events can provide an opportunity to renegotiate. Likewise, if rail infrastructure is private or owned by a public company with a certain degree of autonomy, any unforeseen capacity enhancement can provide an opportunity to renegotiate access charges.

1.1.1.3 Bounded rationality

As pointed out by Hicks (1935), "The best of all monopoly profits is a quiet life." This quest for a quiet life and the bounded rationality of the people in charge of various kinds of decision, together with the lack of competitive pressure, can lead to a sort of slack resulting mainly in the need for too

many inputs (employees, energy, capital inputs and so on). Inefficiency is the direct result of the non-optimising behaviour associated with the enjoyable situation of monopoly.

For infrastructure, this lack of optimising behaviour can be observed in the main tasks of infrastructure management: maintenance and slot allocation (for railways), with a resulting decrease in the capacity offered and an increase in maintenance costs.

1.1.1.4 Externalities

Externalities are a well-known cause of market failure. Transport infrastructure is a source of positive and negative externalities. Regulation is necessary in this case because positive externalities tend to be produced in insufficient quantity and negative externalities tend to be produced in excessive quantities. One of the main problems may be pollution. As public transport is less polluting than private vehicle usage, an inter-mode arbitration may be necessary, with some incentives (regulation or taxation). Urban road pricing (e.g. in the UK and Norway) provides an example of this kind of incentives scheme, destined to reduce the modal share of the most polluting mode. Another example is provided by some railway infrastructure charging systems (e.g. in France and Sweden), with lower charges for freight trains in order to reduce road transport's negative externalities. Low access charges are supposed to induce a modal shift toward rail transport, considered as less polluting. The effectiveness of the infrastructure charging system in reducing pollution is, of course, dependant on the efficiency of the freight operator.

Network externalities can be found where the whole system creates more value than the mere addition of its parts. It is obvious that infrastructure (road and rail) benefits from network economies. But it mostly implies that the design of infrastructure should be such that good interconnections link the different parts of the network. The main issue, therefore, is the division of a network into different sub-networks and the necessary co-ordination between them, which may concern infrastructure design, maintenance and operation.

1.1.1.5 Public policy objectives

Infrastructure construction and operation can be undertaken with several redistribution objectives. Particularly, it could be considered as a policy objective to provide infrastructure (highways or railways) in proximity to every city of a certain size. In that case, the design of the infrastructure network is oriented toward regional development rather than economic efficiency. Among possible effects on the operation of infrastructure is, for example, a certain amount of cross-subsidisation. To that end, regulation is necessary.

Any investment, particularly in infrastructure, is grounded in an inter-temporal arbitration. Investing consists in trading present consumption for future consumption. Future consumption is usually compared to present consumption with the help of a discount rate. Governments may have a different discount rate than private households, because governments tend to value the future more highly, compared to private agents. Thus, the State might tend to intervene in the choices regarding infrastructure investment and operation because of its concern about the future. Private companies are more short-term oriented, especially if they are listed. The stock value becomes a short-term objective, to the detriment of the long-term perspective needed for developing infrastructure. The turbulent history of Railtrack in the UK may serve as an example of this conflict between long-term and short-term objectives.

1.1.2 Why regulation is problematic

We will address more precisely the options for the regulator in the last part of this chapter. But it is noteworthy that the regulator aims at correcting the market failures previously presented. The objectives may be both income redistribution and efficiency, but we will concentrate on efficiency.

1.1.2.1 Multi-product firms

Many firms subject to regulation produce a variety of products. Furthermore, many firms produce different ranges of goods or services which do not have exactly the same characteristics. This concerns both the demand and production sides: for example, a car producer may produce different types of car. Very frequently, the producer discriminates by charging a higher price on a certain category of product and accepting a lower margin on another. Moreover, very frequently some costs are not easily attributable to one kind of product or another.

Sometimes, discrimination occurs as cross-subsidisation: some clients of firms are subsidising others. This is frequently the case with utilities. Regarding transport infrastructure, if there is an access charge, it is highly possible that, with the charging system, certain parts of the network are actually subsidising other parts.

Thus, regulation of multi-product firms may be more problematic because it is difficult to:

- know the cost of each product or activity;
- know what is the degree of cross-subsidisation;
- appreciate the extent to which this possible cross-subsidisation is consistent with the objectives of the regulator.

1.1.2.2 Capture theory

It is common that interest groups try to influence the regulator in order to promote their interests. This tendency can evolve into a more damaging form: the capture of the regulatory body (Stigler, 1971, Becker, 1983). By the word "capture", the capture theory suggests that the firm(s) or some other stakeholders can strongly influence the regulatory body, even to the point where the regulation would be turned against the theoretical objectives of regulation. For example, the regulatory body would protect firms from competition instead of promoting competition. Another example concerns the cross-subsidisation mentioned before. Some groups of customer may capture the regulatory body to benefit from lower prices or better quality.

1.1.2.3 Asymmetric information

Another basic problem with regulation is that the regulator lacks information about the firm regulated. For example, in past times in France, it was not uncommon for some state-owned companies to take the precaution of hiding strategic information. We can wonder whether this may still be the case today. This asymmetric information concerns both the firms (the costs and efforts necessary to reduce them) and their environment (the demand and external factors influencing costs or demand). Usually, two kinds of informational constraints are distinguished:

- Hidden actions, or moral hazard, undertaken inside the firm, for example, to reduce costs;
- Hidden information on the firm's identity (its ability to exert effort), or adverse selection, known by the firm regulated but not by the regulator.

1.1.2.4 Other regulatory constraints

To design a regulation scheme, Laffont and Tirole (1993) add two more constraints to the informational constraint:

a) Transactional constraints: the costly constraints of writing and enforcing contracts;
b) Administrative and political constraints: respecting the rules governing administration and politics.

1.2. The options for the regulators

This part of the report owes much to Laffont and Tirole (1993) and to Carlton and Perloff (2000). We will successively present various forms of regulation or deregulation, finishing by yardstick competition. A special emphasis will be given to price regulation.

1.2.1 Government ownership versus privatisation

At a very high level of generality, according to Vickers and Yarrow (1988), there is not very much to say about privatisation, apart from the fact that "ownership matters". But the effects of government or private ownership depend on the market structure, the type of regulation and incentives and, broadly speaking, the overall environment of the firm.

Regarding infrastructure, it seems difficult to privatise rail infrastructure, as the British experiment has shown. This does not mean that private, integrated companies cannot be efficient. On the contrary, the US or Japanese examples illustrate the efficiency of private, vertically-integrated railways. But it seems doubtful that a private firm can deal with infrastructure alone, because infrastructure development may not be consistent with its objectives, which include, among other things, the maximisation of the value of the shares. Nonetheless, if infrastructure is state-owned, its management can be private. This may be the case, for example, for motorways, which are government-owned but which may have private management.

1.2.2 Price regulation

Price regulation consists in directly fixing the price of the monopoly. Numerous theories have addressed this question: we will only briefly present the main approaches with respect to their relevance for infrastructure and their "power". Laffont and Tirole (1993) define the power of an incentive mechanism by its ability to induce the firm to reduce its costs.

1.2.2.1 Cost plus versus *price cap*

In a cost plus contract, the firm does not bear any of its costs. The regulator pays to the firm an amount of money corresponding to the average costs, including a "fair" profit. This incentive scheme has very low power. On the contrary, with a price cap contract, the prices are limited by an index usually related to inflation and often including some anticipated productivity gains. This kind of contract is extremely powerful, as the firm keeps the benefits of its effort to reduce costs. During the first years of the British railways reform, Railtrack's access charges were predominantly regulated by price capping. This led to cost reductions, as predicted by the theory, but the lack of incentive to invest in capacity enhancement and safety gave rise to their well-known problems.

As a matter of fact, any price cap regulation is periodically revised to take into account the evolution of the costs of the firm regulated. Hence, opportunistically, the regulator may benefit from the cost reduction policy of the firm. It follows that the firm may not have a strong incentive to reduce costs. This is called the "ratchet effect".

Apparently, there is a trade-off between an incentive regime which tends to extract the informational rent[1] but which is almost totally non-cost-reduction-incentive (cost plus) and a regime which leaves at least part of the informational rent to the regulated firm but which provides a strong incentive (price cap). In the real world, the regulatory regime very often lies somewhere between the two polar cases.

Regarding infrastructure, the pricing system is very often debated in terms of marginal cost *versus* average cost. The cost-reduction incentive questions are somewhat neglected in favour of questions of welfare maximisation, considering that the costs are given.

1.2.2.2 Short-run marginal cost, long-run marginal cost or average cost pricing?

Since 1844 and the famous paper from Dupuit, the marginal cost has been favoured for infrastructure charging. But the question is still debated. Average cost pricing still has some supporters. One of the most appealing arguments in favour of average cost, pointed out by Coase (1945), is as follows. Charging to marginal cost gives us no information on the utility to produce the whole infrastructure. In other words, depending on the shape of the demand curve, we do not know whether the infrastructure is socially desirable or not, even if the willingness to pay for it covers the marginal cost.

Nevertheless, it is generally agreed that the short-run marginal cost (hereafter SRMC) constitutes the best theoretical solution to the question of infrastructure charging. The principle is rather simple: SRMC enables the running of all vehicles for which the additional costs borne by society are inferior to the utility of the vehicle for society, supposedly represented by the willingness to pay. Even if the principle is simple, the question of what kind of costs have to be included in the SRMC is rather tricky. In particular, environmental and congestion costs lead to specific problems. But SRMC may provide no incentive to invest, as congestion is a source of revenue. To remove this disincentive to invest, one might be tempted to charge according to other principles, among them, long-run marginal costs (hereafter LRMC).

LRMC is defined as the cost of an additional vehicle when the infrastructure is optimally adapted to demand. Another approach, comparable to a certain extent, is simply to charge the long-run average incremental cost of expanding capacity where capacity is scarce. The main difficulty with these approaches is to practically calculate the amount to be charged. Increasing the capacity of an infrastructure segment leads to the question of indivisibilities. Thus, the cost may vary considerably from place to place. On the contrary, this charging system leads to more stable charges over time and thus facilitates the establishment of contracts between operators and infrastructure managers (Nash *et al.*, 1999). Long, stable contracts may justify specific investments such as rolling stock. There are some arguments, therefore, in favour of long-run marginal cost pricing, even if it would deprive the public of valuable services (the services with prices lying between short-run marginal cost and long-run marginal cost). But one of the major drawbacks of marginal costs (short-run or long-run) stems from the fact that marginal cost pricing does not fully cover the costs. Thus, to cover the non-allocated part of the full costs, some public funds may be needed. They are not given for free or, in other words, there is a cost when obtaining public funds. So, other pricing methods may be used in order to adjust the level of public subsidies to the desired target level.

1.2.2.3 The Ramsey principle

The Ramsey principle aims at differentiating charges according to the value attributed to the services. Again, the principle is quite simple: those services able to pay something above the short-run marginal cost must not be eliminated, because they contribute towards covering the total cost and they deliver a socially desirable service. The services with high value can pay higher prices. The idea is thus to set up a tariff based on the demand for the services offered. The higher the value placed by an operator (and its consumers), the more the demand will be inelastic. So, in order to have a higher tariff for the operator who values the services the most, the tariff must be inversely related to the elasticity of demand for that service. This is called the inverse elasticity rule. Mathematically, the mark-up over marginal cost is proportional to the inverse of the price elasticity of the demand.

Ramsey pricing provides a useful theoretical guideline. However, it requires a great deal of information. Both marginal cost and elasticity of demand must be quantified with a certain degree of accuracy. The Ramsey rule has also been criticised for its failure to protect captive customers.

There are other pricing schemes, but a full analysis of infrastructure charging would be outside the scope of this report. Nonetheless, one conclusion may already be outlined. Regarding infrastructure, pricing is seen as a tool to reach given objectives and not necessarily as an instrument to achieve cost efficiency.

1.2.3 Rate-of-return regulation

Instead of directly regulating prices, regulators can use the rate-of-return (hereafter ROR) regulation to control the rate of return to capital used in utilities. This kind of regulation allows the regulated firm to achieve a rate of return on the fixed capital, once the depreciation of the capital stock and operational expenses have been taken into account.

Thus, ROR is defined by:

$$ROR = \frac{pQ - OE - D(K)}{p_K K}$$

where:

Q is the quantity of product (or service) sold at a price p;
OE represents the operating expenses;
D(K) is the depreciation of the productive capital;
p_K is the average price of capital goods.

The main drawback of ROR regulation is the tendency for the regulated firm to overinvest in capital (Averch and Johnson, 1962). Thus, the common wisdom about ROR regulation is that it is inefficient, although it may lead to lowering prices to under the monopolistic level. However, regulators in the US are no longer in favour of ROR regulation, although it may be practised in other countries.

1.2.4 Making a market contestable

The theory of the contestable market appeared in the US in the late 70s and early 80s (Baumol, Panzar and Willig, 1982). The main idea is that, if a monopoly practices monopoly pricing or is inefficient, the threat of a new entrant can compel him to improve efficiency and reduce prices. Thus, in a market where the incumbent monopoly can be contested, the monopoly rent is reduced to zero. Making a market contestable consists mainly of reducing the sunk costs and the entry barriers. But this theory is grounded on a number of assumptions, which are not very frequently verified, notably the credible threat from entry (which means low sunk costs and no strategic game protecting the incumbent). The contestable market theory is even more difficult to analyse for multi-product firms and complex price schemes. We will not expand on this important theory because it seems to us inappropriate for infrastructure.

If the infrastructure manager is also the owner of the track, obviously it would not be welfare-improving to duplicate infrastructure, even if the new infrastructure management is more efficient.

If the infrastructure manager is not the owner, several criteria contribute to considering the threat of a new entrant as damaging: the infrastructure manager must develop some long-term contracts with both clients and suppliers. The threat of a new entrant may make the establishment of contracts uneasy, to the extent that they can then be renegotiated by the new entrant.

1.2.5 Franchising

Theoretical works on franchising or "competition for the market" mostly derive from the work of Demsetz (1968). What we call franchising here is recurrent competitive tendering to allocate the whole market to a single firm for a given period of time. Thus franchising is also called competition for the market, as opposed to competition in the market. A contract is established between the successful bidder and the public body in charge of the tendering process. Regarding transport, franchising already has a long history, with the urban transport concessions system in France and the railway operators in the reformed UK rail system.

Franchising certainly appears to provide a way to put local monopolies under pressure, and this kind of regulation is becoming more and more frequent in Europe (Van de Velde, 2003). Much has been written on franchising – see, for example, Preston and Nash (1993) for a theoretical analysis of railways franchising and White (2000) for an analysis of competitive tendering for London buses. The main difficulties in implementing franchising are the following:

- The possible collusion between would-be bidders;
- The existence of sunk costs, which can hinder true competition;
- The establishment of contracts, which is costly and, of course, incomplete. This leads to many problems, in which opportunism can play a role.

Regarding infrastructure, we will assume that franchising is not very efficient to regulate infrastructure management, but to make this point, we will need some developments on the infrastructure management activity itself and thus we will address this question in the fourth part of the report. Let us remark that the franchises for infrastructure are generally quite long and thus cannot lead to the introduction of strong competitive pressure. This is the case for the French motorways, the London Underground and the future French-Spanish high-speed line, Perpignan-Figueiras.

1.2.6 Yardstick competition

1.2.6.1 What do we call "yardstick competition"?

There is no precise definition of what yardstick competition is, given that the associated theory has led to various methods of implementation, as described in Chapter 3. However, we can distinguish two main meanings given to the term "yardstick competition".

- On the one hand, this expression refers to a regulatory framework, based on comparisons. It is a virtual form of competition between similar regulated firms, as in Shleifer's proposal (see next page). It consists in estimating what should be the best prices and subsidies, by comparing the performances of various regulated firms. The regulator, by setting the correct prices and subsidies, can lead the firms to produce an effort which increases welfare.

- On the other hand, yardstick competition refers to the basic and relatively informal use of comparisons by a regulator who wants to improve expertise and reduce the informational asymmetry he faces. In that sense, yardstick competition is an additional expert tool used by the regulator to improve the efficiency of another regulatory framework (franchising, for example).

1.2.6.2 Yardstick versus benchmarking

Benchmarking is sometimes presented as a kind of yardstick competition. We do not consider benchmarking in this report because, for us, it is undertaken by firms in order to improve their processing or methods, and not by the regulator.

1.3. Conclusions

Infrastructure provision has to be regulated, notably because it leads to a monopoly and thus to market power. Other market failures are also present (opportunism, bounded rationality and externalities). Infrastructure regulation is also problematic, mainly because of asymmetric information.

We will consider in this report that the aim of the regulator is cost efficiency.

Among the options offered to the regulators, we would like to stress the following points:

1. Government ownership seems desirable for an infrastructure separated from operation;
2. Price regulation is submitted to various objectives, but not primarily the cost efficiency of the infrastructure provider;
3. Rate-of-return regulation should play a limited role, notably because of the Averch-Johnson effect;
4. Making the market contestable as well as franchising constitute options of limited scope because of the long-term effect of many contracts and many maintenance operations. This point will be developed in the fourth chapter of the report.
5. Yardstick competition might be included in the regulatory framework of infrastructure management, but before going further, we will discuss the theoretical foundations of yardstick competition. This is the subject of the next chapter.

2. SURVEY OF THE MAIN THEORETICAL APPROACHES

In this part of the report, we detail yardstick competition mechanisms through a review of the main theoretical approaches. First, we present the principle of yardstick competition as defined by Shleifer and the ways of implementing it. In the next two sections we analyse the informational benefit of the comparisons in the two cases of informational constraints (moral hazard and adverse selection). Then we discuss the main limits to yardstick competition in a static context (correcting external heterogeneity) and in a dynamic one (investment incentives and collusion).

2.1. Principle of yardstick competition

In this section, we present generalities about introducing yardstick mechanisms. First, we discuss Shleifer's model, then we analyse how yardstick competition can be implemented.

2.1.1 Shleifer's model

Shleifer (1985) is the originator of theoretical mechanisms of yardstick competition. The model, inspired by Medicare's observations (see Chapter 3), defines the concept of yardstick competition relatively simply.

2.1.1.1 Model and assumptions

Shleifer considers $N \geq 2$ similar firms, operating on geographically separated but identical markets and producing the same output. He assumes that the firms face the same demand curve. Each firm, i, is characterised by its marginal cost, c_i, and its investment in cost reduction, $R(c_i)$. For each firm i, the regulator sets the price, p_i, and distributes a subsidy (a lump-sum transfer), T_i. In order to set prices and subsidies according to firms' performances, the regulator compares each firm to its defined yardstick, as described in the following paragraph.

2.1.1.2 The yardstick

For each firm i, consider:

$$\bar{c}_i = \frac{1}{N-1}\sum_{j \neq i} c_j$$

$$\bar{R}_i = \frac{1}{N-1}\sum_{j \neq i} R\left(c_j\right)$$

Each firm i is assigned its own "shadow firm" which serves as the benchmark in yardstick competition. Shleifer shows that the regulator can achieve the economic optimum by setting:

- the price of firm i, equal to the average marginal cost of the other firms: $p_i = \overline{c}_i$
- the lump-sum transfer to firm i, equal to the average investment in cost reduction of the other firms: $T_i = \overline{R}_i$

Using game theory, he shows that every firm's optimal strategy is to reveal its true cost, c_i, and investment in cost reduction, $R(c_i)$. Thus each firm i is forced to compete with its yardstick, defined by the performances of the other firms. This scheme implicitly defines the costs that firm i has to reach and the investments in cost reduction it should make in order to have a positive profit.

2.1.2 Implementation of yardstick competition

Before reviewing applications of yardstick competition to utilities regulation (next chapter), we ask three main theoretical questions related to the implementation of such a regulation framework. We also show that there are various ways of implementing yardstick competition.

2.1.2.1 Correcting heterogeneity

Through the simple model of Shleifer, we noticed that it is crucial to correct the external heterogeneity as, in fact, the assumption of identical environment is not acceptable. The costs of firms are also influenced by factors associated with each environment. In fact, in the same article, Shleifer presents a simple model to correct external heterogeneity. As each yardstick is defined by the costs and investments of the other firms, uncorrected heterogeneity would bias the yardstick and then lead to an inefficient mechanism. This important question will be discussed hereafter.

2.1.2.2 Yardstick competition and franchising

Franchising, also called competition for the market, has been presented in Chapter 1. It appears that the concepts of yardstick competition and competition for the market are clearly different. However, we suggest that these two ways of regulating firms are, in fact, quite complementary.

On the one hand, competition for the market is efficient to the extent that the regulator has sufficient information to prevent possible collusion between bidders[2]. It is possible to prevent collusion by reducing its benefit (which consists of a monopoly rent shared by the cartel members). This means that, if this profit can be sufficiently diminished, as is possible using yardstick competition, the regulated firms would have fewer incentives to collude. So yardstick competition can contribute to the successful implementation of competition for the market.

On the other hand, consider a firm -- regulated through yardstick competition -- which seems to be inefficient, according to comparisons. The inefficiency of that firm can be due to mismanagement or to the contract negotiated with the regulator. Therefore, it is possible to reduce this inefficiency by changing the management of the regulated firm; hence, the regulator should be able to manage the firm's exit from the market. However, given the specificity of transport infrastructure services, such an exit must not lead to service interruption. This is the reason why a recurrent bidding process (such as franchising) seems to be an interesting solution.

Thus, competition for the market and yardstick competition are complementary and, to a certain extent, enforced by each other.

2.1.2.3 How to develop incentives?

Shleifer's model and other theoretical models use a financial mechanism linked to the results of the comparisons to enforce competitive pressure. Those models show that the expectation of gains and the threat of penalties generate such competitive behaviour. However, this is not the only way to reach this goal. In particular, the use of comparisons in addition to franchising induces competition between the regulated firms, for the two following reasons:

- A reputation effect arises: no firm can afford to have a much worse image than the others. This effect appears as soon as the regulator widely spreads the results of its comparisons;

- A belief effect comes into play: if the regulator can convince its firms of its attachment to the results of the comparative mechanism, he can create additional competitive pressure. In particular, the threat of non-renewal of the contract at the next bidding can induce such an effect.

So, yardstick competition can be implemented through different ways: in connection with franchising or not, with a financial mechanism or not. In the two following sections we discuss some more sophisticated models, which illustrate how the comparisons lead to reductions in informational asymmetry.

2.2. Informational benefit of the comparisons in the "moral hazard" case

2.2.1 What is "moral hazard"?

The moral hazard problem arises from an informational asymmetry which the regulator faces, concerning the regulated firm's behaviour. Because the regulator cannot control the firm's actions all the time, the regulated firm tends to adopt opportunistic behaviour. This is the very reason why the regulator has to provide effort incentives for its regulated firm, as explained in Chapter 1. However, the firm's production depends not only on the effort it has exerted, but also on hazard (influence of a third party, the weather, the overall economic situation, etc.). Firms can face a commercial risk (if demand decreases) and an industrial one (if costs increase more than prices). Due to this hazard, the firm faces uncertainty which may or may not be cause for fear:

- Consider a state-owned firm, operating a public utility service which suffers no interruption. This firm is quite sure that it cannot go bankrupt because of its public ownership and the necessity of ensuring continuity of the service. It is assured of being recapitalised sooner or later. Thus, such a firm does not fear uncertainty, knowing it is potentially insured against risk. Such behaviour towards risk is known as "risk-neutral".

- Consider, now, a private firm. It cannot assume that its shareholders will constantly accept the financial impacts of random events. Hence, such a firm fears uncertainty. It is said to be risk-adverse. In a regulatory framework, the regulator has to pay such a firm a fixed incentive (a risk insurance) in order to offset the risk which the firm faces.

We now analyse, in this moral hazard case, how comparisons may reduce the uncertainty and thus the risk insurance payment.

2.2.2 Modelling uncertainty

Many authors model the uncertainty which affects firm i's production by the following elements:

- A common uncertainty parameter, η, which affects all firms in the sector. This variable mainly includes the situation of the sector (overall economic, social, political, competitive situations and so on);

- An independent, particular risk, ε_i, which includes localised impacts on production (weather, difficulties caused by a third party…). The ε_j's are assumed to be independent and normally distributed.

Thus, the risks $(\eta+\varepsilon_j)$ they are facing are correlated to the extent of the similarity between the regulated firms. Mathematically, this would be translated as :

$var(\eta)>>var(\varepsilon_j)$.

2.2.3 Theoretical results

Holmström (1982) shows that, in such a configuration, the use of comparisons improves welfare. The results are the following:

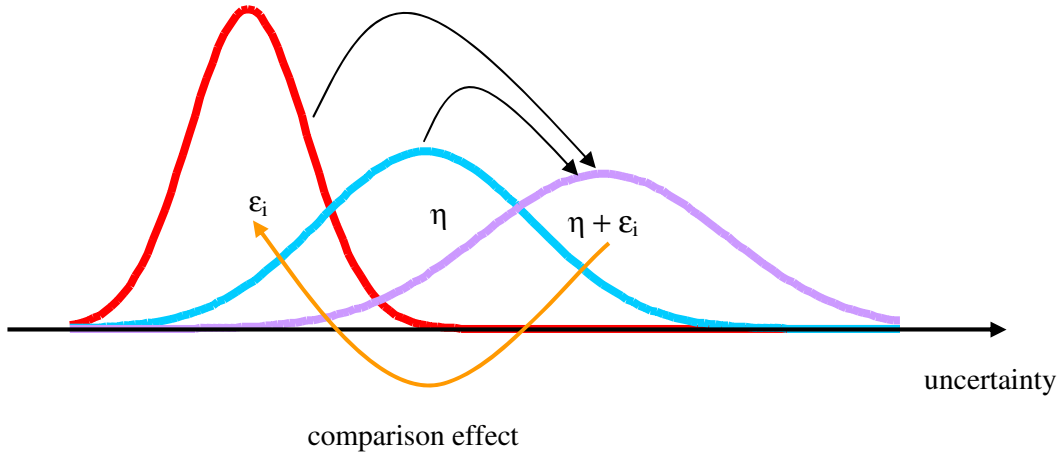
- If firms are risk-neutral, yardstick competition is as interesting for the regulator as the other regulation schemes which do not use comparisons;

- If firms are risk-adverse, the economic efficiency is improved by comparisons as soon as there is common uncertainty ($\eta\neq0$). This improvement increases with the number of firms compared.

2.2.3.1 Reduction of uncertainty

The underlying idea is somewhat intuitive (see Figure 1). Capturing all the relevant information about η, included within the outcome measures, leads the regulator to neutralise the impact of common uncertainty. The outcomes then depend only on the ε_j's and the efficiency of the firms (assuming that external heterogeneity is corrected). Thus, the risk a firm faces is reduced from $(\eta+\varepsilon_i)$ to ε_i.

Figure 1. **The comparisons reduce uncertainty**



Following this result, the higher the ratio, var(η)/var(ε_j), the more the comparisons are relevant. The regulator who reduces uncertainty in this way can then decrease its risk-insurance payment to the regulated firms.

2.2.3.2 Application to transport infrastructure services

We believe that transport infrastructure services are scarcely concerned by this mechanism, due to the relatively low uncertainty which the regulated firms face. Var(η) is low because transport infrastructure service firms operate in a very stable environment: η mainly takes into account changes in labour legislation and price increases, which are not a major source of uncertainty. Var(ε_j) is not so low, and includes mainly climatic events (which can cause severe and costly damage to infrastructure) and uncertainty related to the geological knowledge of the subsoil. However, with transport infrastructure services it is relatively easy to determine the financial impact due only to those random events. This enables the regulator to insure, *ex post,* its regulated firm against risk.

Furthermore, transport infrastructure service firms are quite risk-neutral because of their public ownership and/or social utility (the service suffers no interruption). Therefore, although yardstick competition cannot have a negative effect on the regulation of transport infrastructure services, its implementation in such a moral hazard case does not seem very relevant. However, infrastructure companies' regulators could be rather more concerned by the adverse selection situation which is addressed in the following section.

2.3. Informational benefit of the comparisons in the adverse selection case

2.3.1 What is adverse selection?

The adverse selection problem arises from the informational asymmetry which the regulator faces concerning the regulated firm's identity (its productivity). For a given production and a given payment by the regulator, consider two types of firm:

- A low-efficiency firm, which needs to exert much effort to provide the service required, and whose income (related to its efforts) is just sufficient to ensure its participation;

- A high-efficiency firm, which needs to exert only minimum effort to provide the same service, and whose income (related to its efforts) ensures it a quiet life. However, such a firm could provide a much better service, exerting much more effort; but it would do so only if it receives sufficiently high incentives. This transfer is also called "informational rent".

In the adverse selection problem, due to informational asymmetry, the regulator cannot distinguish between the high-efficiency and low-efficiency firms. Therefore it cannot, without cost, force high-efficiency firms to exert greater efforts, instead of leading a quiet life. To prevent those firms from passing themselves off as low-efficiency firms, the regulator has to:

- encourage them to reveal their high-efficiency identity; to reach this goal, it has to pay them the informational rent;

- discourage them from adopting a low-efficiency identity; to achieve this objective, it has to reduce the attraction of the low-efficiency firms' situation (by decreasing the level of service required, and thus their income); this production distortion leads the regulator to reduce the informational rent which it has to pay.

Therefore, in order to optimise welfare, the regulator must propose to the firms a revealing menu of contracts. This menu should include one contract (low production and low income) chosen by the low-efficiency firms and another contract (high production and high income) for the high-efficiency firms. We now analyse, in the case of adverse selection, how comparisons could reduce the informational asymmetry regarding the identity of firms, and hence the informational rent and production distortions.

2.3.2 Modelling productivity

The firms' productivity model is designed like an uncertainty model: one distinguishes a common element (whose impact would be reduced by the comparisons) and a residual, specific one. Therefore, many authors model the productivity parameter, β^i, which characterises each firm i (assumed to be risk-neutral) by the following elements:

- A common part, $b \in \{b_1, b_2\}$, which is the same for every similar firm in the market. This variable reflects statutory, regulatory, organisational etc., particularities of the market: b_1 corresponds to high-efficiency firms, whereas b_2 refers to low-efficiency ones.

- A specific part, ε^i, which is assumed to be independent and identically distributed in a given interval. This individual characteristic includes the effects on the efficiency of firm i's decisions (investments, management, etc.).

2.3.3 Theoretical results

Auriol (2000) proves that, in such a configuration, the use of comparisons improves the regulation efficiency (for more details, see Annex). She considers that the regulator proposes to firms a revealing menu of contracts. Using game theory, Auriol shows that firms are encouraged to choose the contract which corresponds to their productivity parameter, β^i. Hence, they reveal their common part, b_1 or b_2. With this information, the regulator is able both to reduce the informational rent of the high-efficiency firms (b_1) and to suppress the distortions on the production of low-efficiency firms (b_2).

2.3.3.1 Reduction of informational rent

Following the above mechanism, the regulator can detect inconsistent announcements (when a firm announces a characteristic corresponding to a common part $b=b_2$, and the others announce characteristics corresponding to $b=b_1$). So the regulator can retaliate against the firm giving such an inconsistent announcement (by excluding it from the market, for example), in order to make the mechanism revealing. Thus, a high-efficiency firm can no longer behave as if it were of low efficiency. The informational rent is then reduced.

2.3.3.2 Suppression of production distortions

Given that b_1 firms can no longer pretend to be of type b_2, it is not necessary for the regulator to maintain the distortions on the b_2 firms' production. Thus, the comparisons lead the regulator to induce those firms to exert the optimal level of effort.

2.4. Correcting external heterogeneity

After having presented Shleifer's model, we noticed the need to correct the external heterogeneity which influences a firm's performance, and thus the yardstick it is compared to. In this section, we illustrate how external heterogeneity impacts on a firm's performance and how it can be corrected. One usually distinguishes the two following types of heterogeneity:

2.4.1 Endogenous heterogeneity

This section refers to Bivand & Szymanski (1997), who analysed the effect of endogenous heterogeneity, which they define as the "spatial dependence effect".

2.4.1.1 What is endogenous heterogeneity

For the situation under review -- transport infrastructure regulation -- endogenous heterogeneity arises as soon as the regulatory framework differs between the firms compared. If yardstick competition is implemented by a unique, centralised regulator, every regulated firm faces the same requirements. In such a case, there is no endogenous heterogeneity. However, if there is decentralisation, each firm faces its own regulator's requirements. Endogenous heterogeneity may arise from the particular policies of each regulator.

For example, a regulator (1) could require a very high quality of service from its firm 1 and pay the (relatively high) associated price. In this case, firm 1 seems inefficient (because of its high costs) if the service quality is not taken into account. Another regulator (2) could be financially benevolent toward its firm 2. This firm would appear relatively inefficient, for its costs are high, due to the

benevolence of its regulator. If comparisons made by a third regulator (3) cannot take into account those particular policies, they would be biased in favour of its firm 3, which would be compared to apparently inefficient firms.

Such policies generate externalities which influence the firms' performances and, subsequently, the comparison yardstick.

2.4.1.2 The case of transport infrastructure services -- how to correct endogenous heterogeneity

Concerning transport infrastructure services, regulation is rarely centralised: on the one hand, local (urban and rail transport) services are often regulated by local authorities. On the other hand, major road infrastructure (highways, civil engineering works) are nationally, but apparently loosely, regulated. In this context of decentralised regulation, the involvement of the different regulators' in the development of their transport infrastructure services would be a major source of endogenous heterogeneity. Contrary to other network industries (water and energy supply, telecommunications) which have to serve an entire population, at all times and in all places, transport infrastructure services are frequently defined by local or national government and decisions are motivated by political considerations.

Such decisions impact on the costs of services and the level of traffic.

- To reduce the impact on costs, it is most important to correct this heterogeneity by including the necessary corresponding factors (quality and/or frequency of service, etc.) in the estimated cost function, as should be done to correct exogenous heterogeneity (see below).

- To reduce the influence on traffic level, the regulator has to base his comparisons on costs which are independent from the level of traffic[3]. For example, an urban transport regulator should compare average costs per bus-kilometre rather than average costs per passenger-kilometre. This depends not only on the firm's performance but also on the fare, the frequency of services, etc., set by the regulator.

2.4.2 Exogenous heterogeneity

For transport infrastructure, exogenous heterogeneity is caused by the environment, rather than the firm or the regulator. Such environmental heterogeneity arises from geographical constraints (relief, flood risk, network spatial structure and so on), demographic differences (density of population), social particularities (delinquency rate) and many other factors.

In order to correct such sources of distortion, the solution consists in introducing a measure of those factors in the estimation of the cost function. This means that the accounting cost of the service has to be corrected through an econometric process before being used in the comparisons. But this method leads to other difficulties, which we briefly expose in the next section.

2.4.3 Choice of a cost function

We have just seen that when correcting heterogeneity it is necessary to specify a cost function which includes the different factors of heterogeneity. However, this solution should be reserved for the comparison of sufficiently similar firms, for several reasons:

- First, it is not easy to include every source of heterogeneity. In particular, the quantitative measurement of some sources may be very difficult;

– Secondly, the mathematical form chosen for the cost function influences the results of the correction.

Thus, the choice of cost function (specification and form) may impact on the heterogeneity correction, which would then be imperfect. This is the very reason why yardstick competition should be applied to similar firms, operating in relatively close environments which do not need a wide heterogeneity correction.

However, concerning infrastructure transport services, the financial impact of the main factors is relatively well known. For example, the maintenance cost of an infrastructure is predictable, given traffic characteristics (number of vehicles, speed, weight, etc.). Hence, it is quite easy to ensure a good correction of the heterogeneity between compared infrastructure transport services.

2.5. Some theoretical limits

Before concluding this chapter, we analyse yardstick competition in a dynamic context since, in fact, the robustness of a regulation framework has to be considered through time. We discuss two common limitations of a regulation framework: the investment incentives, which have to reach an adequate level (neither under-investment nor over-investment) and collusion, which we have already mentioned concerning franchising.

2.5.1 Reaching an adequate investment level

2.5.1.1 Spillovers and under-investment

Dalen (1998) shows that yardstick competition could reduce investment incentives in some cases. Taking Auriol's model framework, he considers that every investment can be split into two parts:

– A "firm-specific" part, which increases only the efficiency of the investing firm, i.e. this investment allows the firm to improve its own productivity (ε^i);

– An "industry-specific" part, which by externality ("spillovers") increases the common efficiency of all the firms in the sector (b).

Of course, yardstick competition promotes firm-specific investments. On the other hand, when investments benefit other firms, yardstick competition reduces investment incentives. This can be particularly the case with research and development, whose outcomes may spill over into other firms at a very low cost, compared to what the first firm has invested.

So it is necessary, before implementing yardstick competition, to evaluate its impact on the research and development activities in the sector. However, we believe that this problem of spillovers and under-investment should not particularly affect transport infrastructure services.

2.5.1.2 Regular long-term investments

In the first chapter, we explained that difficulties could arise because private firms are rather short-term oriented. Another theoretical investment problem can then be foreseen when yardstick competition is applied. There is a risk that excessive competitive pressure on the current expenditure

of a regulated firm will lead it to delay or reduce long-term investments. This problem, which particularly affects infrastructure services, is due to the difficulties of long-term regulation. We will discuss this in the last chapter.

2.5.2 The collusion between compared firms

2.5.2.1 What is collusion?

Collusion is usually defined as the co-operative behaviour between firms which should, on the contrary, compete against each other. Such firms can agree to form a cartel and to co-ordinate their prices or information, in order to reduce the competitiveness of the oligopolistic market. This leads them to benefit from higher (monopolistic) prices and to increase their profit at the consumers' expense. Hence, the resulting allocation of resources is sub-optimal. This is why collusion has to be avoided.

Note that a collusive agreement requires the participation of every firm operating on the associated market. If one firm does not collude, it can practice lower prices than the collusive ones, and hence it gets the whole market and the cartel fails. Thus, it is difficult for the colluding firms to preserve the cartel from any deviation, given that each of them has opportunistic incentives to deviate.

Moreover, collusion between firms could occur even in a non co-operative context (see Tirole, 1993). This tacit collusion arises from the threat of a vigorous price war, after one firm cuts prices. This provides strong incentives to refrain from actively competing, so that a relatively quiet life can be maintained on the market.

2.5.2.2 The reasons for which collusion is likely to appear

The threat of collusion arises when firms are regulated by yardstick competition because they can foresee that they will obtain zero rent from the mechanism proposed by the regulator if they play non-co-operatively. Hence, they are willing to co-ordinate their messages to counteract the regulator's power.

In particular, the "revealing principle" used by Auriol may become inefficient if firms collude. Consider the model of firms' productivity in the case of adverse selection, with a parameter $b \in \{b_1, b_2\}$, distinguishing between high- and low-efficiency firms. At first sight, the high-efficiency firms seem able to collude and announce a low-efficiency characteristic. This would allow them to exert a lower than optimal effort. In this case, collusion, directly distorting the comparison yardstick, reduces the incentive effect of yardstick competition.

However, Auriol's particular model is so designed that the mechanism suppresses any collusion incentive. The best choice for the firms, whatever their productivity characteristic, consists in choosing the contract which corresponds to their productivity.

2.5.2.3 Why the risk of collusion is not always so high

In practice, the theoretical threat of collusion has not been observed in the above-mentioned sectors. Many reasons can be proposed:

- The high number of firms compared greatly limits the development of collusion. This is due to the weak strength of collusive agreements when the number of participants increases, as it becomes difficult to involve each of them in the cartel and prevent deviation.

- The comparisons made by the regulator lead it to easily detect any deviation from competitive towards collusive behaviour.

- Moreover, comparisons based on accounting data (expenses, investments, etc.) reduce the ability of the regulated firm to hide or modify their information.

For these reasons, we believe that collusion is not such an important risk when firms are regulated by yardstick competition.

2.6. Conclusions on the main theoretical approaches of yardstick competition

To conclude this chapter, we can note that yardstick competition seems to be an interesting way of regulating similar monopolies. The use of comparisons provides strategic information for the regulator. In particular, it reduces the informational asymmetry which the regulator faces. Some mechanisms may hinder or at least limit collusion. The theory is flexible enough to give rise to various kinds of application. This is what we will observe in the next chapter.

3. APPLICATIONS OF YARDSTICK COMPETITION TO UTILITIES REGULATION

In this chapter, we present the main applications of yardstick competition to the regulation of various utilities (hospitals, water utilities and so on), with special attention to transport services (the Japanese railways industry and Norwegian buses). The objective of this chapter is to show the many different ways of implementing a regulation scheme which is related to some kind of yardstick competition.

3.1. The activity-based reimbursement of hospitals

Historically, the reimbursement of hospitals is the first application of yardstick competition. The American programme Medicare began in 1983, two years before Shleifer's theoretical proposal. In the beginning, this regulation scheme was very close to pure yardstick competition.

3.1.1 Activity-based reimbursement and yardstick competition

3.1.1.1 Principles

Given the inefficiencies of hospital systems in the OECD countries (high costs or long queues), many countries have reformed them, using activity-based financing[4]. However, both the objectives and the implementation of this financing system vary greatly between the different countries. The common point of these activity-based systems is the definition of so-called "Diagnosis-related Groups" (hereafter DRG). A DRG is a group of illnesses or pathologies which need the same hospital treatment. Hence, it is possible to define and compare the costs of each hospital for every DRG.

3.1.1.2 Objectives

The three main objectives of an activity-based healthcare system are the following:

- To fit the financing to the activity;
- To encourage the hospitals to reduce cost increases;
- To promote competition in order to reduce rents.

The top objective of each government depends on the key inefficiency observed. On the one hand, countries facing over-high costs (the US, Germany, France) would be interested by cost growth reduction. On the other hand, countries facing over-long queuing (the United Kingdom, Scandinavia) would prefer promoting competition in order to encourage hospitals to improve their productivity.

3.1.2 Implementation and effects

It appears that the implementation of an activity-based system depends on the top objective followed by government.

Governments aiming at reducing the costs of their hospital system use price regulation, according to Shleifer's proposal. However, hospital reimbursement systems are rather mixed: they are not uniquely related to the comparisons. For example, Medicare, the American health insurance for old and handicapped people, is based on such comparisons. Since 1983, Diagnosis-related Groups have been defined and, for each of them, the reimbursement price is based on the average of the costs (corresponding to this DRG) observed in the hospitals. Of course, external heterogeneity is corrected by including demographic and geographical data in the comparisons.

Although it is difficult to evaluate the efficiency of a price regulation, it seems that the effects of activity-based reimbursement are positive: lengths of hospital stays decreased, growth in costs was reduced, while the quality of healthcare was preserved.

Other governments, whose objectives are to reduce hospital queues, promote competition through comparison. For several years, the British Government has compared the costs of its hospitals for each DRG. However, it seems that competitive pressure related to the use and spread of the comparisons was not strong enough to achieve productivity increases for hospitals.

3.2. Water utilities in Great Britain

This case study is interesting, because it illustrates the benefits of the relatively informal use of comparisons by a regulator.

3.2.1 The regulatory framework and the comparisons

The 26 Welsh and English water utilities are regulated by the Office of Water Services (OFWAT). Although firms compete for the market, the regulator has to exert pressure on them, given the very long duration of the contracts (25 years). Prices are regulated through a price cap mechanism (see Chapter 1). The initial level of price limit and its evolution (the anticipated productivity gains) are set by OFWAT every five years. Moreover, the regulator compares the firms' performances every year, in order to promote "comparative competition". Performance measures are based on service quality, accounting and financial data, and are used as exposed in the next subsection.

3.2.2 How are the outcomes of comparisons used?

As explained in the previous chapter, such an informal use of comparisons leads to a competitive effect -- due to a reputation effect -- and to an increase in the regulator's expertise.

3.2.2.1 The reputation effect

The comparisons lead to a reputation effect which encourages firms to behave competitively, because no-one can afford to have a worse image than the others. This effect arises due to the publication of the outcomes of the comparisons. OFWAT issues a yearly report showing the firms' results concerning their cost efficiency. Moreover, the report includes tables of data related to service quality, so that customers can evaluate the performance of their water utility. This has proved to exert considerable competitive pressure on the firms.

3.2.2.2 An increase in the regulator's expertise

Comparisons also lead the regulator to increase its expertise concerning two points. First, OFWAT uses the comparisons to set the parameters of the price cap scheme. A price ceiling and its evolution are defined for each firm given the outcomes of the comparisons, among other data. Secondly, the comparisons lead the regulator to detect certain inefficiencies (for instance, too many leaks in a given network) and then to reduce the informational asymmetry which he faced at the beginning.

It has been shown[5] that the improvement in the firms' efficiency was more significant for the firms which were inefficient at the beginning of the regulatory period. This proves that, in addition to the price cap, comparisons have played a role in the competitive pressure exerted on the firms by OFWAT.

3.3. Dual sourcing

Dual sourcing consists in introducing a second firm into a monopolistic market. It sometimes appears that the advantages of a duopoly structure can be preferred to the duplication of fixed costs, given the difficulties of regulating a monopoly. This is particularly the case when the fixed costs are due to large investments in research and development. The two following examples show that dual sourcing offers the project manager a mixture of co-operation and competition, with various advantages.

3.3.1 US Defence Department

The American Defence Department has been using dual sourcing since 1980, in particular for important and expensive systems[6]. For instance, the Advanced Medium Range Air-to-Air Missile, the Tomahawk Cruise Missile and F15 fighter aircraft engines were produced by two competing firms following the dual sourcing approach.

The process is as follows. After competition at the design stage, one firm is selected by the Department of Defence to develop its proposal and to undertake the initial production. After initial production, the Department of Defence can transfer technology to a second firm and force the two potential producers to compete for a production contract. Technology is transferred through a "learning buy" in which the second firm receives generous compensation for delivering a few products.

The benefits of this dual-sourcing practice are the following:

- The reduction in the expected marginal cost due to the use of yardstick competition between the two producers;

- Avoiding capture of the Department of Defence by a sole firm with a large informational and learning advantage over its potential competitors. One of the objectives of dual sourcing is to share knowledge, so that competition can occur for further development and/or production contracts.

3.3.2 The Petrona Twin Towers of Kuala Lumpur

The Petrona Towers of Kuala Lumpur are the tallest buildings in the world. Each tower was built by a different firm: Samsung and Mitsubishi.

The project manager decided to use dual sourcing for the following reasons:

- With the technical difficulties involved in building such tall towers, some uncertainty was foreseen concerning problems which could arise needing rapid solutions. So dual sourcing allowed both contractors to learn from each other when one of them faced a difficulty.

- Having two separate construction teams developed a healthy competition between both parties to see who would do a better job. For instance, Mitsubishi had begun the second tower one month after Samsung, yet finished its construction at the same time. In this case, dual sourcing provided incentives to compete.

3.4. Transportation services

The two following cases of regulation frameworks related to yardstick competition are interesting for two main reasons. Firstly, because they concern transport services (Japanese railways and Norwegian buses) and, secondly, because the use of comparison is not informal as in the two previous cases. Conversely, comparisons are used in the financial regulation mechanism. This is why we develop in more detail the underlying yardstick mechanism for both cases.

3.4.1 Yardstick competition implemented in the Japanese railway industry

This part of the report owes much to a recent article (Masaru Okabe, 2004) and would not have been possible without the help of Makoto Ito and Fumitoshi Mizutani. The yardstick competition system in Japan is original and seems at least partly successful (Mizutani, 1997). The railway companies subject to yardstick competition are the following:

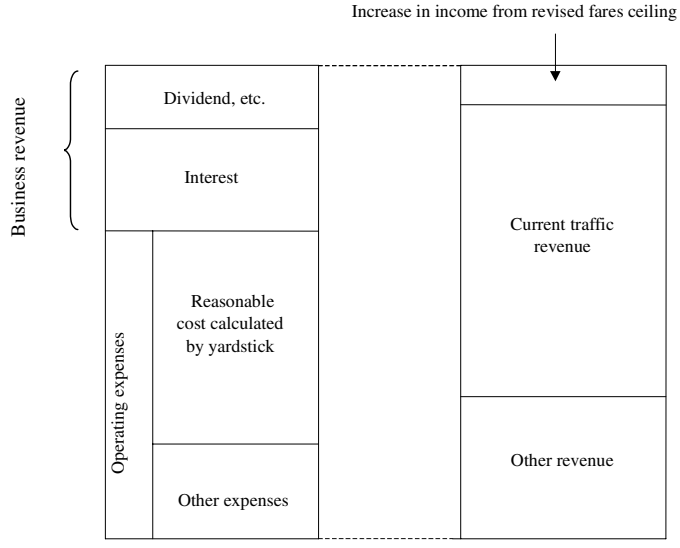
- The 15 major private companies;
- The six Japan Railways (the JRs, the six regional companies created in 1987 when JNR was broken up into six passenger companies bound for privatisation and a freight company);
- Ten public subway companies.

With those 31 companies, yardstick competition is now applied to railways carrying about 95 per cent of rail passengers in Japan. We successively present a general overview of the system, the yardstick competition mechanism and a brief summary of the assessment made by Mizutani.

3.4.1.1 General overview of the incentive scheme

The regulatory scheme is a combination of price cap, ROR and yardstick competition. The fares must be under a ceiling (they can be significantly under this ceiling). The ceiling is calculated based on a "reasonable cost", which is the result of yardstick competition. Thus, a comparatively cost-efficient company can have a greater "business revenue". But there is a trade-off between revenue and lower fares, as this procedure allows companies to change fares, under the ceiling, by simple notification to the Government. However, to avoid unfair parallel competition with cross-subsidisation, some fare reductions are submitted for government approval.

Figure 2. **Overview of the incentive scheme in Japanese railways**



Source: adapted from Okabe, 2004.

The reasonable cost is calculated by a yardstick competition process, which will be described below; the total cost is determined by ratios. The procedure depends on the kind of firm regulated, but this implies a fair profit on capital invested. The ceiling is given by the total cost, including the reasonable cost calculated by the yardstick.

3.4.1.2 The yardstick competition

There are five cost categories (see Table 1). For each of those costs, the unit cost is calculated by dividing the costs by the number of facilities (associated volume).

Table 1. **Facilities and explanatory variables for each kind of cost**

Costs	Facilities (associated volumes)	Main explanatory variables
Tracks	Track length	Rolling stock km per track km (Log)
Catenary	Catenary length	Electric multiple unit km per catenary km
Rolling stock	Number of rolling stock	Rolling stock km per rolling stock
Train operation	Route-kilometre	Train km per route km (Log)
Stations	Number of stations	Number of passengers per station

Source: Okabe, 2004.

Then for each cost category, a model is calibrated with explanatory variables intended to be correlated with the unit cost. The main explanatory variables are given in Table 1, but the model may differ according to the class of railway. For track costs (mainly maintenance), the details are given in Table 2.

Table 2. **Calculation of the standard unit cost for maintenance cost (maintenance costs per kilometre of line)**

Group of railways	Regressors
15 large private rail companies	Tunnels and bridges percentage Log (traffic density)
6 JR rail companies	Log (traffic density) Amount of snowfall
10 public subway companies	Log (traffic density)

With these models, it is possible to calculate the base unit cost for each category of costs and for each railway company. This may be interpreted as the unit cost the railway should arrive at if this company's average cost is as calculated by the model.

The standard cost, for each category of cost, is then calculated for each company by multiplying the base unit cost by the actual associated volume.

Finally, the total of the "yardstick" costs can be calculated; this is called the reasonable cost.

There are then two possibilities:

- If the actual cost is inferior to the reasonable cost, half of the difference is added to the actual cost and this becomes the new reasonable cost. This corresponds to sharing the benefit of higher productivity between the consumer and the producer.

- If the actual cost is higher than the reasonable cost, the reasonable cost is not changed. This means that the company will have to increase productivity in order to reach equilibrium.

There is another incentive mechanism. If a company decreases its productivity, as measured by the difference between actual cost and reasonable cost, the reasonable cost is diminished by half of this difference. On the other hand, in the case of cost efficiency improvement, the reasonable cost is increased by half of the improvement.

The total cost and the ceiling are calculated as previously indicated.

3.4.1.3 The effect of yardstick competition in Japan

The regulation system described here is the result of a reform which took place in 1997. The only published analysis in academic journals is the one by Mizutani (1997), which was carried out before the reform. Nonetheless, some points are noteworthy:

- Large private operators subjected to yardstick competition have made significant efforts to improve cost efficiency;

- The effect of yardstick competition is not only to reduce operating costs but also to reduce the variance of operating costs, as would a "true" competitive effect;

- However, measured against a reference model, managerial efficiency improved for only 40 per cent of operators.

It would be very interesting to have new measures of the effect of yardstick competition under the new regime, particularly including the six JRs.

3.4.2 Norwegian buses

This section owes much to Dalen and Gomez-Lobo's recent paper (2003). In Norway, responsibility for local bus transport is decentralised to the 19 regional governments (counties). Each county is free to choose its own regulatory policy; many of them have already implemented yardstick competition for years. Yardstick competition is used by some regional regulators to determine the level of subsidy granted to the companies (there is a large number of bus companies in each county).

The level of subsidy granted to the n^{th} company, S_n, is equal to the difference between expected traffic revenue (R_n) and expected costs (C_n):

$$S_n = R_n - C_n = P_n y^e_n q_n - \sum_k c_{k,n} - c_{adm,n}$$

- Concerning the expression of the expected revenue, P_n is the average fare level set by the regulator for the bus services of company n, y_n^e is the expected number of passengers per kilometre produced and q_n is the total number of kilometres produced, defined by the regulator.

- Concerning the expected costs, C_n is the sum of the k-inputs (drivers, fuel, maintenance, bus capital) costs, the $c_{k,n}$, and the administration costs, $c_{adm,n}$, set proportional to $\sum_{k \neq fuel} c_{k,n}$.

Each input cost $c_{k,n}$ of company n is defined as follows: $c_{k,n} = p_k \sum_i \sum_j a_i^j q_{i,n}^j$, where p_k is the yardstick unit (per kilometre produced) price for input k, $\sum_i \sum_j a_i^j q_{i,n}^j$ is the sum of the kilometres produced by company n, weighted by the route type, i, and the bus type, j, (coefficients a_i^j). This weighted sum corresponds to the correction of heterogeneity due to route type and bus type.

So, yardstick competition applys because the subsidy granted to a company is related to the benchmark prices of the k-inputs the firm should buy to produce its transportation service.

Dalen and Gomez-Lobo (2003) prove that bus companies regulated under yardstick competition become more efficient and reduce their costs faster than the others; in this case, yardstick competition provides more dynamic incentives for cost reduction by operators.

3.5. Conclusions on the applications of yardstick competition to utilities regulation

Through various case studies, we have illustrated how regulation schemes could be more or less related to yardstick competition. This proves an interesting property of yardstick mechanisms: their flexibility, in the sense that they can be used in multiple cases, given various constraints and objectives. Moreover, yardstick competition seems particularly interesting for some kinds of transport services regulation. We now discuss the particular case of transport infrastructure services.

4. YARDSTICK COMPETITION IN INFRASTRUCTURE PROVISION

In this final chapter, we will address the question of implementing yardstick competition for infrastructure provision. To this end, we will first analyse the infrastructure provision and see that it consists of a set of several different activities. We will then consider step by step the different activities undertaken by an infrastructure provider. This part of the report principally concerns railways, but some points will also be relevant for tolled motorways.

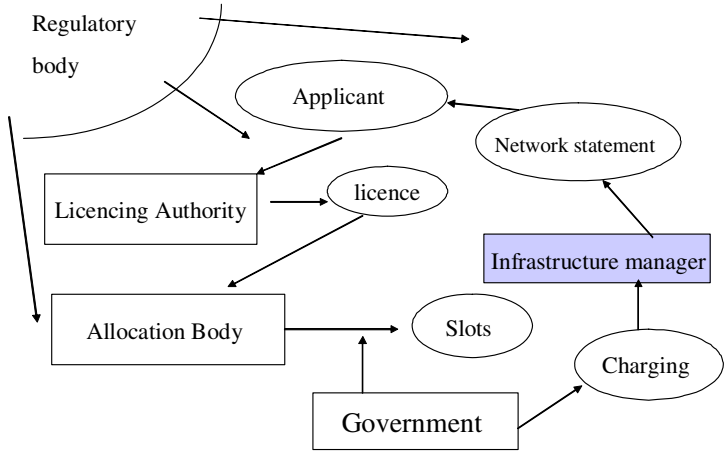
4.1. Analysis of infrastructure providers' activities

Even if infrastructure is frequently owned by the State, it does not imply that their managers should escape from all competitive pressure. Before discussing the applicability of yardstick competition to the various activities of the infrastructure manager, it is useful to analyse more closely what are the given tasks for an infrastructure manager. This will allow us to define the term "infrastructure manager".

Basically, infrastructure management consists in providing transport capacity at certain prices and certain costs. To provide capacity, infrastructure has to be built, maintained and "operated". Maintenance can be further divided into light maintenance and renewal. All these activities are undertaken under the supervision of what we will call an infrastructure manager. The activities of an infrastructure manager may be extremely diverse. However, the initial investment decision may not be under his responsibility. The authorities may themselves set up new infrastructure and contribute to investment before giving responsibility of operation to the infrastructure manager. Furthermore, maintenance tasks may or may not be carried out by the infrastructure manager.

Moreover, regarding railways, infrastructure operations are part of a complex system. The access system, as derived from the European directives, is shown in Figure 3.

Figure 3. **The institutional context of European Infrastructure Manager (railways)**



It follows that a certain number of activities may or may not be under the sole responsibility of the infrastructure manager: notably, slot allocations and infrastructure charging. Thus the traffic on one or other of the lines may not be the result of his management. Hence, the actual traffic on the infrastructure is to be considered as an exogenous factor and any evaluation of the pricing system should be made without taking into account the rules governing the infrastructure manager's activities.

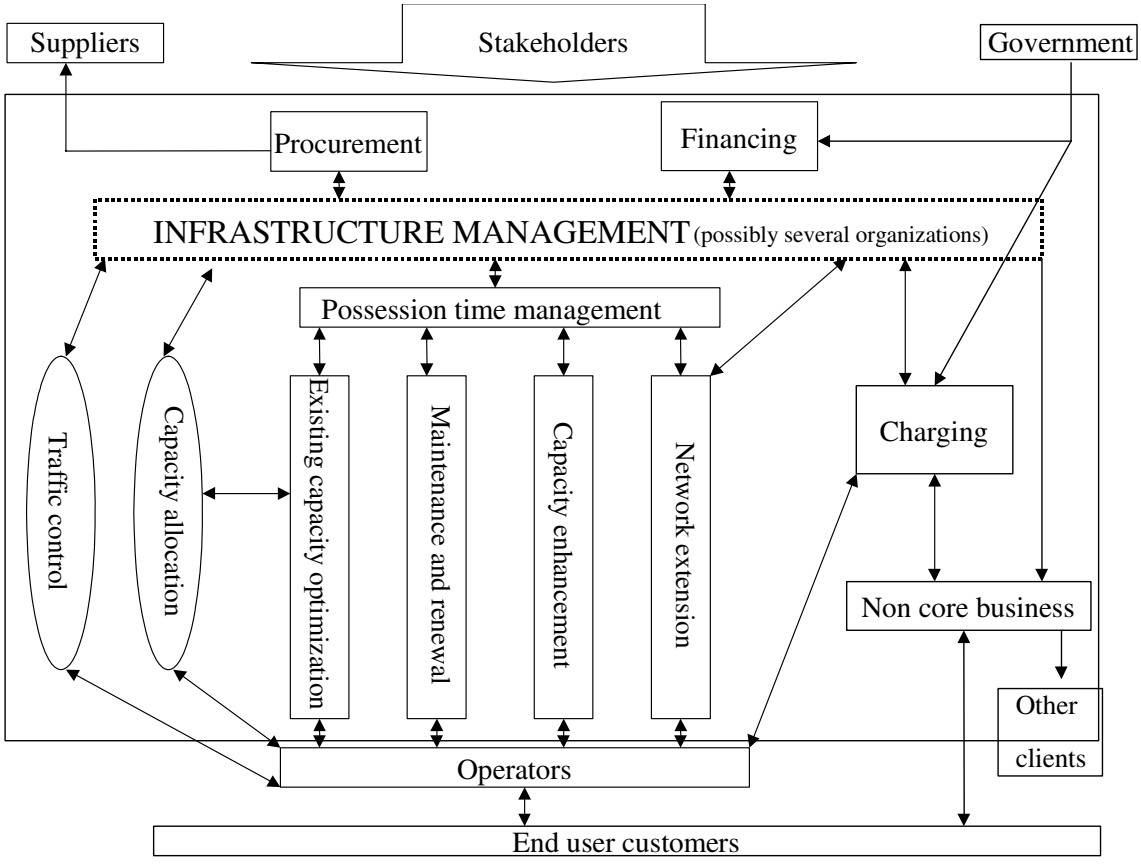
To a certain extent, we can say that infrastructure managers have a multiple production function:

- Firstly, producing infrastructure, i.e. a physical object defined by its material characteristics;
- Secondly, producing capacity, i.e. the quantity and quality of slots which can be offered on the infrastructure;
- Thirdly, the final "product", i.e. the circulation of trains.

This only applies to railways. For roads, only the first production is relevant, and the final product is capacity, partly the result of the initial design of the highways, and thus not fully under the responsibility of the infrastructure manager.

Regarding railways, the links between the different tasks and the various stakeholders are represented below in the form of a "core process map" (Figure 4).

Figure 4. **Core process map of "infrastructure management"**



Source: Improverail research project for the European Commission.

Given that slot allocation and traffic control may be confided to bodies independent of the infrastructure manager, as well as the fact that charging may be decided by authorities with diverse objectives, the primary functions of the infrastructure manager can be resumed as:

- optimisation of existing capacity;
- maintenance and renewal;
- capacity enhancement; and
- network extension.

All these processes are linked with so-called "possession time", i.e. the time necessary for maintenance. Maintenance can be dissociated from other activities because its costs are easily observable.

So, in order to see to what extent yardstick competition can be implemented, we will consider here successively the three steps: infrastructure construction; maintenance; and other infrastructure operations. Finally, we will add a few remarks on the "production of traffic", because many railways are integrated. We will not discuss here the question of vertical separation, which is outside the scope of this report. However, we can observe that vertical separation facilitates the application of yardstick competition for maintenance, because the necessary data are more easily observable. Nonetheless, we will consider the application of yardstick competition to infrastructure management by integrated firms.

4.2. Initial construction

Regarding initial construction, the same analysis is valid for road and rail infrastructure. The usual way to introduce competitive pressure on the suppliers is through competitive tendering. Could yardstick competition be jointly used with competitive tendering? The main difficulty with yardstick competition is that there is a very large exogenous heterogeneity of civil engineering works (engineering structures).

Yardstick competition could be useful for certain simple civil engineering works, such as the construction of roads or rail track over flat countryside. We will not expand on this, however, given the fact that yardstick competition may have only a limited role to play in putting the suppliers of initial construction under pressure. It may be that what is called "dual sourcing" could be usefully implemented, if the infrastructure has two comparable parts.

4.3. Maintenance

Regarding maintenance and operation, we need to distinguish between road infrastructure on the one hand and rail infrastructure on the other. Rail infrastructure in the EU Member States should be considered under a different angle, because their management has to be separated from train operations. In the European Union, the infrastructure manager does not provide a final product, given that its customers are the railway companies (the "railway undertakings" in the official terminology of the European institutions) who operate trains. It follows that maintenance of rail infrastructure can be:

- subcontracted to independent companies;
- undertaken by the main railway company (the "historic operator" in the European context);
- undertaken by the infrastructure manager;
- a combination of the above possibilities.

For the countries where the railways are vertically integrated (the US, Japan, etc.), the only choice is to rely more or less on outsourcing for maintenance.

During the first years following the British Railways reform, maintenance was largely undertaken by independent maintenance companies. The tendency now is for the infrastructure manager (Network Rail, formerly Railtrack) to reintegrate the maintenance. In France, the maintenance of the track is undertaken by the historic operator, under the responsibility of the infrastructure manager. So far, this historic operator is the only licensed French railways undertaking.

In the case of outsourcing of maintenance works, there appears a clear possibility to implement yardstick competition, the infrastructure manager being the authority and the maintenance companies being considered as regulated. Yardstick competition can therefore apply to:

- the different infrastructure managers;
- the infrastructure manager's suppliers.

This corresponds to different levels:

- the cost of elementary maintenance operations;
- the relevance and the cost/efficiency of the maintenance policy.

We will not expand on the first level: yardstick competition may be used jointly with competitive tendering in order to minimise the cost of external works and to compare it to works carried out by the infrastructure manager itself or by the historic operator. Regarding the question of the possible integration of maintenance companies, we are facing the trade-off between the benefits of integration and the benefits of competition. The transaction costs argue, in principle, for integration, but the price to pay is that there is no more competitive pressure.

Regarding the second level, which consists in the comparison of maintenance costs, the question is, in short: how to judge that a maintenance policy is appropriate? Several remarks can be offered:

1. Maintenance has to be suited to the kind of traffic on every segment of line. It follows that the set of lines under the responsibility of the infrastructure manager has to be broken down into different categories, notably according to the number of tonnes transported per year.

2. Maintenance consists of two kinds:

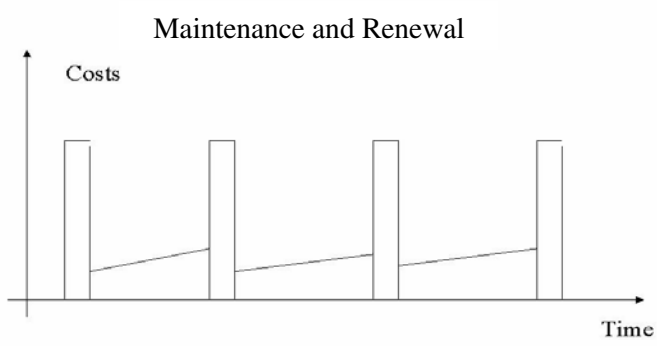
   - light maintenance;
   - renewal, which consists of heavy works undertaken to obtain a quasi new infrastructure.

So, ideally, the maintenance policy has to be compared against the life cycle cost, which includes Maintenance and Renewal (see Figure 5).

Light maintenance is carried out during a period, hereafter called the possession time, in which no train can run. For some obvious reasons, the maintenance costs decrease when the possession time increases. The fact is that the opportunity cost of the possession time tends to vary greatly, according to the traffic on the line and in particular according to the degree of congestion. If the line is operated with sufficient free capacity, the opportunity cost of the time devoted to maintenance is zero.

The last remark establishes a link between maintenance and what we call infrastructure operation. This means that for heavily loaded lines the assessment of the maintenance policy's relevance cannot be isolated from the general traffic management.

Figure 5. **Simplified representation of the life-cycle costs**

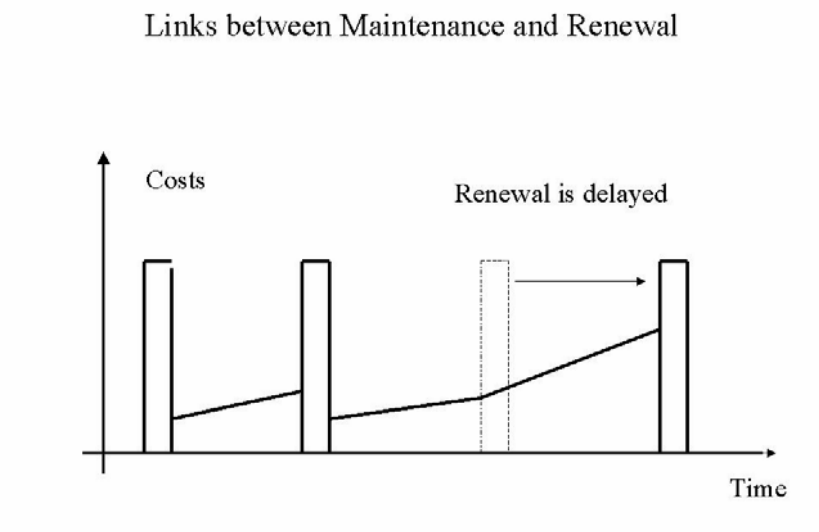


What are the conclusions from the previous remarks concerning yardstick competition for maintenance?

First, the traffic, probably measured in tonnes, has to be taken into account and thus included in the external factors which might affect the maintenance cost. Whether the dependency is purely linear or more complex has to be studied.

Secondly, any yardstick comparison should take into account the life-cycle cost and not only the current maintenance cost. This is not an easy task. A first solution could be to have a dummy for light maintenance and renewal and to compare uniquely what is comparable. But these two kinds of costs are linked. If the infrastructure manager delays renewal for a significant time, the maintenance cost increases (Figure 6).

Figure 6. **The absence of renewal increases the maintenance costs**



Another solution, more complex and needing more data, would be to update the costs, by the classical methods, over the total life cycle.

If the rail networks to be compared are big enough and if there are sufficient numbers of networks taking part in the yardstick competition, the problem of life-cycle costs may be less disturbing. During an average year, some parts of each network will be renewed and others will be only lightly maintained. From the average over several years, the cost efficiency of the maintenance policy can be compared. To avoid unjustified subsidy variations, a moving average can be used. Another possibility to avoid this problem would be to admit that some infrastructure managers will bear renewal costs periodically and that equilibrium will be reached over the entire life cycle.

The third remark has radical implications. Even if the possession time is presented as an external factor which tends to introduce external heterogeneity into the comparison of maintenance policy, if we consider maintenance costs alone, we are able to capture only part of the infrastructure manager's performance. A low-cost maintenance due to an excessively large possession time would be cost-effective but may not be welfare-improving because of the opportunity cost of the lost capacity. We do not encounter this problem with the Japanese yardstick competition mechanism. As the companies are integrated, they tend to internally optimise the trade-off between large possession time and more capacity offered to the trains. If the line is not saturated, the opportunity cost of the possession time is zero. However, for the saturated parts of the network, this leads us to the next step: capacity management.

4.4. Capacity management for railway infrastructure

We now address the ability to perform certain tasks which leads the infrastructure manager, possibly associated with other public bodies, to offer capacity. These tasks are:

- Optimisation of capacity to maximise the production of slots;
- Enhancement of capacity, including elimination of bottlenecks if desirable;
- Slot allocations;
- Traffic control.

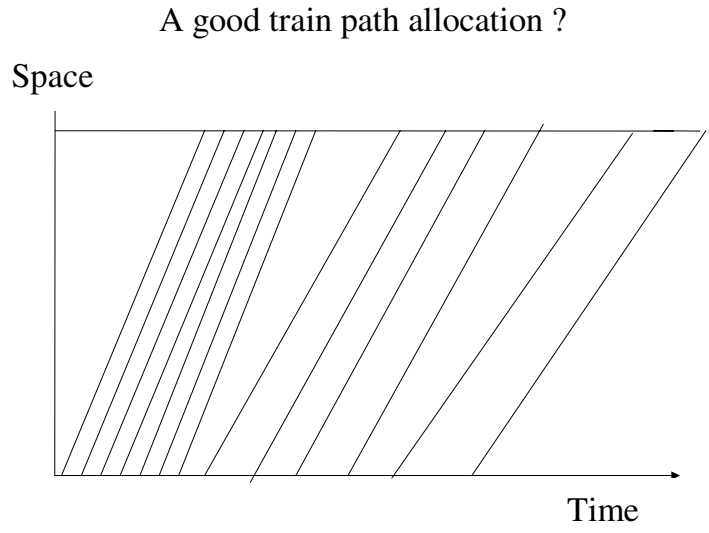
Implementing yardstick competition for capacity management would be very difficult; nonetheless, it is possible to make some comparisons which give a small taste of this sort of competition. Before tackling this question, we will look at the main aspects.

First, we are facing diverse outputs: the capacity itself is an output as well as the number of trains running on the network, but this traffic is not the result of the infrastructure manager's efforts, as the pricing system is out of the manager's domain. Another output is safety.

An additional, important difficulty is the heterogeneity of the train path. A freight train is generally slower, and a passenger through-train is naturally less capacity-consuming than a slow train.

Last, but not least, the possibility to group together the slow and fast trains can increase capacity. In other words, the actual capacity, measured by the number of trains which can run on the line per day, is not an objective *per se*, as the timetable has to be suited to the demand (see Figure 7).

Figure 7. **Grouping of similar trains increases capacity, but slot allocation has to take demand into account**



We might add that the intervals between the trains depend partly on the infrastructure, but they also depend on the technology of each train. So, even if the actual number of trains running on a line is, to a certain extent, the result of the infrastructure manager's effort, it is greatly exogenous.

Theoretically, it should be possible to implement yardstick competition on a set of very similar lines, with similar trains -- for example, regional lines, with no interference from freight trains during the daily service. But the unit to be submitted to comparison is the line, and a large part of management costs are common to the different lines which are parts of the same network. So, the common costs, not clearly attributable to any line in particular, make any cost comparison between the lines rather difficult.

If the lines are saturated, it is quite impossible to value the opportunity cost of the slots wasted due to mismanagement. So we propose to ignore this kind of cost, even if some comparisons can lead to the view that an apparently saturated line can actually bear more traffic. This would be a qualitative assessment based on comparison, but as there is no clear incentive mechanism, we will not consider this comparison as yardstick competition.

If we now consider different networks, the infrastructure manager's accounts can be shown as follows:

Subsidies = (Capital costs) + Maintenance costs + Management costs - Infrastructure charges

It is necessary to estimate if management costs can be broken down into two components:

- a component linked to the length of tracks;
- a component linked to the traffic.

For the most part, these are labour force costs.

To implement yardstick competition, it is desirable to exclude capital costs and infrastructure charges.

For capacity management costs, some further studies are necessary, but most likely the unit cost should be calculated by the ratio of costs divided by the number of train-kilometres, controlled for length of track and composition of traffic, and broken down according to the type of traffic -- unsaturated regional lines, intercity, high-speed or suburban lines, etc. This yardstick competition for management costs has to be implemented to compare infrastructure managers which perform similar tasks.

4.5. The production of transport services

First, it is necessary to point out that this paragraph does not concern road infrastructure. Moreover, for the infrastructure manager, traffic is to be considered as given and exogenous, i.e. the traffic actually on the infrastructure is not under the sole responsibility of the infrastructure provider. Indeed, we can wonder whether it is possible to implement any kind of yardstick competition for the infrastructure provision of integrated railway companies. For sure, production, as usually measured in terms of traffic units (tonne-kilometre plus passenger-kilometre or any other linear combination of the two) is too linked to external factors to provide a reliable indicator of the effort exerted by companies. It is probably possible to compare different traffic evolutions and try to assess to what extent those evolutions are due to external or internal factors. Nevertheless, any assessment of the performance of railways on the basis of traffic could be strongly biased (Savignat and Nash, 1999). So it would be preferable to make the hypothesis that train-kilometres and vehicle-kilometres are the actual outputs of the railway companies.

For the integrated systems, it is possible to implement yardstick competition, as the Japanese example clearly illustrates. But can we implement yardstick competition with integrated companies on the maintenance costs alone and can we use vertically separated and vertically integrated infrastructure companies together in the same sample?

If there is accountancy separation (as is the rule in the EU), the maintenance department of the integrated company can be treated as an infrastructure company. There might be another system of yardstick competition for the operating companies or the operating divisions of the integrated companies. If there is no accountancy separation between infrastructure maintenance and train operation, it seems difficult to implement yardstick competition. Some internal cost allocation rules can hinder fair competition, thus the competition has to be between integrated companies. The example of Japan illustrates that such competition is possible and is apparently improving efficiency.

Why should yardstick competition on infrastructure alone be desirable? More generally, why should it be desirable to separate infrastructure costs from the costs of operation?

- First, to estimate access pricing, it is important that the maintenance costs are well identified, because infrastructure charges almost always incorporate marginal costs, of which maintenance costs are a part;

- Second, it is important to compare different maintenance policies in order to assess their cost efficiency (for example, integrated maintenance *versus* outsourcing);

- Third, the time scale of maintenance is different from that of train operation; hence it is relevant to compare maintenance and operations separately;

– Finally, a good knowledge of optimal maintenance costs is necessary to know whether an infrastructure should be shut down or not.

To conclude on the production of traffic, we will offer two final remarks:

1. When measuring the efficiency of firms, the production of traffic is not an output to be considered, even if good traffic management can increase traffic.

2. When possible, it is desirable to compare the efficiency of the maintenance policies of different firms, whether integrated or not.

4.6. Conclusions on yardstick competition in infrastructure provision

Infrastructure provision consists of the following activities:

– Initial construction;
– Maintenance;
– Capacity management;
– Production of traffic.

The first conclusion is that there appear to be limited possibilities to implement yardstick competition for initial construction, notably because of the high heterogeneity of engineering works.

Yardstick competition can be implemented for maintenance works and could be useful for the maintenance divisions of integrated companies.

Yardstick competition for capacity management can be implemented on similar networks to detect wasted capacity, but a more flexible kind of comparison can also produce some efficiency gains.

Finally, this analysis of infrastructure management leads us to the conclusion that franchising may not be very efficient due to life-cycle costs. There are some risks of opportunistic behaviour through diminishing maintenance costs during the franchise period, to the detriment of another franchisee who then bears higher renewal costs.

5. SUMMARY AND CONCLUSIONS

Regulation is necessary for infrastructure because of the important market failures described in Chapter 1. By its very nature, infrastructure is a monopoly. Moreover, infrastructure management gives ample room for opportunism and could be qualified as being subject to bounded rationality. Furthermore, the transport industry gives rise to numerous externalities.

Among the various options offered to the regulator, yardstick competition is particularly efficient. The main reasons for this are the following:

- First, it can be used in combination with other tools, among them notably:

  1. Franchising, with the possibility to compare various franchised companies;
  2. Price capping, with the possibility to set a price ceiling with reference to a cost average, as in the Japanese system.

- Second, this regulation mechanism is powerful in the sense that it provides strong incentives to reduce costs;

- Third, yardstick competition has the power to significantly reduce asymmetric information;

- Fourth, it also seems that the risks of collusion can be contained.

Moreover, yardstick competition has proved to be efficient under a number of circumstances in various industries: health care, transport, water supply and civil engineering. However, to the best of our knowledge, it has not been used in infrastructure management. But the Japanese yardstick competition mechanism includes an infrastructure cost comparison. It is also noteworthy to mention that in the Japanese railways industry yardstick competition is based on published and observable data. This may help to significantly reduce the risks of capture.

From the previous analysis of the infrastructure manager's activities, we can conclude that:

1. There is a limited possibility to implement yardstick competition for the initial construction of transport infrastructure, the best way to introduce competition appearing to be competitive tendering;

2. Maintenance can be regulated by yardstick competition; integrated firms can be included if their maintenance accounts are clearly separated;

3. Regarding capacity management, even if some comparisons can help to find productivity gains in slot allocations, it seems difficult to implement pure yardstick competition because of the strong heterogeneity of traffic running on the different lines. However, it appears that yardstick competition could possibly be implemented on very similar lines.

Finally, we can conclude that yardstick competition may become one of the major tools of infrastructure provision regulation. It can be used as part of a more complex framework, including price cap, cost plus or even ROR regulation.

ANNEX: AURIOL'S MODEL

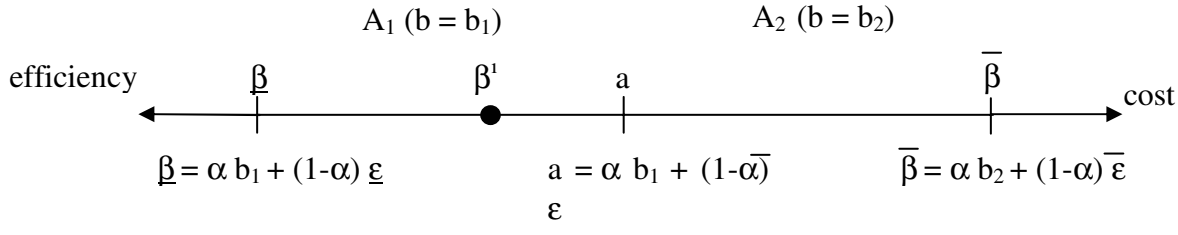
Modelling productivity

We have already explained that the productivity parameter, β^i, consists of the two following elements:

- A common part, $b \in \{b_1, b_2\}$, which is the same for every similar firm in the market. This variable reflects statutory, regulatory, organisational etc., particularities of the market. b_1 corresponds to high-efficiency firms, whereas b_2 refers to low-efficiency ones;

- A specific part, ε^i, which is assumed to be independent and identically distributed over a given interval. This individual characteristic includes effects on the efficiency of firm i's decisions (investments, management and so on).

Introducing the degree correlation between the firms, $\alpha \in [0,1]$, firm i's characteristic is given by: $\beta^i = \alpha b + (1-\alpha)\,\varepsilon^i$ (the more the firms are correlated – α close to 1 –, the more the common part b is important within the characteristic). A high β^i corresponds to high cost, i.e. an inefficient firm. The authors usually assume that $\alpha b_1 + (1-\alpha)\,\overline{\varepsilon} = \alpha b_2 + (1-\alpha)\,\underline{\varepsilon} = a$, which ensures that both intervals $A_1 = [\underline{\beta}, a]$ and $A_2 = [a, \overline{\beta}]$ are disjoint[7] (see figure below).



Theoretical results

Auriol (2000) proves that, in such a configuration, the use of comparisons improves the regulation efficiency. She considers N firms, each of which has a cost function defined by $C^i = \beta^i - e^i$, where β^i is the productivity parameter described above and e^i is the effort exerted by the firm. The regulator proposes a revealing menu of contracts to the firms, defining the transfer Ti to firm i by the following:

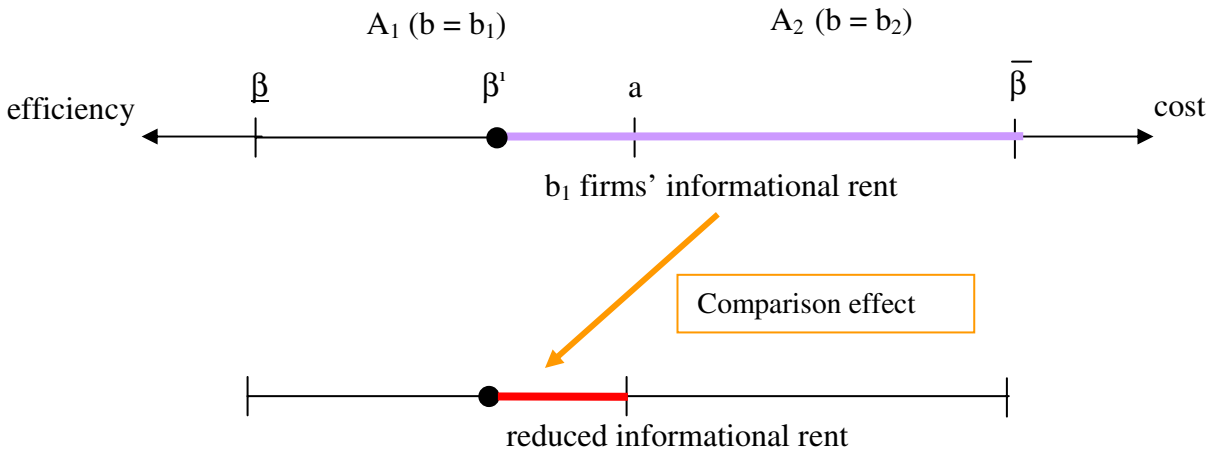
$$T^i = a(\varepsilon^i) + \left[b(\varepsilon^i) C^i + (1 - b(\varepsilon^i)) \frac{\sum_{j \neq i} C^j}{N-1} \right]$$

where: $a(\varepsilon^i)$ is a fixed part, including the informational rent and the effort incentives;
$b(\varepsilon^i) \in [0,1]$ is the insurance against the risk that C^i exceeds the average cost.

The functions $a(\varepsilon^i)$ and $b(\varepsilon^i)$ are so designed that, whatever $b \in \{b_1, b_2\}$ is, $a(\varepsilon^i)$ decreases with ε^i (the more the firms are efficient, the higher the informational rent and effort incentives) and $b(\varepsilon^i)$ increases with ε^i (the more the firms are efficient, the more they have to compete with their shadow firm). Using game theory, Auriol shows that firms are encouraged to choose the contract which corresponds to their productivity parameter, β^i. Hence, they reveal their common part, b_1 or b_2. Knowing this information, the regulator is able both to reduce the informational rent of the high-efficiency firms (b_1), and to suppress the distortions on the low-efficiency firms' (b_2) production.

Reduction of informational rent

Following the above mechanism, the regulator can detect inconsistent announcements (when a firm announces a characteristic in A_2, corresponding to a common part $b = b_2$, whereas the others announce characteristics in A_1). So, the regulator can retaliate against the firm giving such an inconsistent announcement (excluding it from the market, for example), in order to make the mechanism revealing. Therefore, a high-efficiency firm cannot behave as if it were a low-efficiency one anymore. The informational rent on the common part, b, is then suppressed. This truncates b_1 firms' informational asymmetry from $[\beta^i, \underline{\beta}]$ to $[\beta^i, a]$. The informational rent is reduced likewise:



Suppression of production distortions

Given that b_1 firms can no longer pretend to be of type b_2, it is not necessary for the regulator to maintain the distortions on b_2 firms' production. Thus, the comparisons lead the regulator to compel those firms to exert the optimal level of effort.

NOTES

1. The informational rent is defined more precisely in Chapter 2. For now, we can express it simply: this is the rent enjoyed by a firm because the regulator does not know its costs.

2. The collusion problem (the co-operative and hence uncompetitive behaviour between the bidders) is discussed more precisely in section 2.5.

3. Nash (2000) developed such an analysis concerning railways : "*Governments […] frequently intervene in the pricing and output decisions of railways. Performance measures for these railways then typically provide information on a mixture of the performance of the management and of the institutional setting in which it operates. For passenger services it is not uncommon for governments to effectively control the timetable as far as the frequency of service on each route […]. In this situation, arguably the government becomes the customer, and the output the railway produces is a certain level of service, rather than transport for a number of people."*

4. See DREES (2002).

5. See OFWAT (1997).

6. See Riordan and Sappington (1989).

7. Auriol (1993) considers a duopoly structure with non-disjoint intervals.

BIBLIOGRAPHY

Auriol, E. (1993), "Monopole ou duopole : l'effet de comparaison", *Annales d'Economie et Statistique*, 31, pp. 1-31.

Auriol, E. (2000), "Concurrence par comparaison : un point de vue normatif", *Revue économique*, 51, pp. 621-634.

Averch, H. and L. Johnson (1962), "Behavior of the Firm under Regulatory Constraint", *American Economic Review*, 52-5, pp. 1053-1069.

Baumol, W., J. Panzar and R. Willig (1982), *Contestable Markets and the Theory of Industry Structure*, Harcourt Brace Janovich, New York.

Becker, G.S. (1983), "A Theory of Competition Among Pressure Groups for Political Influence", *Quarterly Journal of Economics*, 98, pp. 371-400.

Bivand, R. and S. Szymanski (1997), "Spatial dependence through local yardstick competition: theory and testing", *Economics Letters*, 55, pp. 257-265.

Bouf, D. and P.Y. Péguy (2001), "Is yardstick competition desirable for western European railways?", *International Journal of Transport Economics*, 28-2, pp. 205-227.

Carlton, D. and M. Perloff (2000), *Modern Industrial Organization*, Addison Wesley Longman.

Coase, R.H. (1945), "Price and Output Policy of State Enterprise", *Economic Journal*, 55, pp. 112-113.

Dalen, D.M. (1998), "Yardstick competition and investment incentives", *Journal of Economics and Management Strategy*, 7-1, pp. 105-126.

Dalen, D.M. and A. Gomez-Lobo (2003), "Yardstick on the road: regulatory contracts and cost efficiency in the Norwegian bus industry", *Transportation*, 30, pp. 371-386.

Demsetz, H. (1968), "Why regulate utilities?", *Journal of Law and Economics*, 11, pp. 55-65.

DREES (2002), « La tarification à la pathologie, les leçons de l'expérience étrangère », *Dossiers solidarité et santé*, Hors série, Juillet 2002, La Documentation Française.

Hicks, J.R. (1935), "Annual survey of economic theory: the theory of monopoly", *Econometrica*, 3, pp. 1-20.

Holmström, B. (1982), "Moral hazard in teams", *The Bell Journal of Economics*, 13, pp. 324-340.

Improverail (2003), Deliverable 4. Research project for the European Commission.

Klein, B., R. Crawford and A. Alchian (1978), "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process", *Journal of Law and Economics*, 21, pp. 297-326.

Laffont, J.J. and J. Tirole (1993), *A theory of incentives in procurement and regulation*, MIT Press.

Mizutani, F. (1997), "Empirical analysis of yardstick competition in the Japanese railway industry", *International Journal of Transport Economics*, 24-3, 367-392.

Nash, C. (2000), "Modelling performance: Rail", in: Hensher D.A. & K.J. Button (eds.), *Handbook of Transport Modelling*, Pergamon, pp. 565-575.

OFWAT (1997), *1996-97 Report on water and sewerage service operating costs and efficiency.*

Okabe, M. (2004), "New passenger railway fares", *Japan Railway & Transport Review*, 37, pp. 4-15.

Preston, J. and C. Nash (1993), "Franchising rail services", in: Anthony Harrison (ed.), *From hierarchy to contract, Policy Journals*, pp. 147-165.

Riordan, M. and D. Sappington (1989), "Second sourcing", *Rand Journal of Economics*, 20-1, pp. 41-58.

Savignat, M.-G. and C. Nash (1999), "The Case for Rail Reform in Europe -- Evidence from Studies of Production Characteristics of the Rail Industry", *International Journal of Transport Economics*, 26(2), pp. 201-217.

Shleifer, A. (1985), "A theory of yardstick competition", *Rand Journal of Economics*, 20, pp. 41-58.

Stigler, G. (1971), "Theory of Economic Regulation", *Bell Journal of Economics and Management Science*, 2, pp. 3-21.

Tirole, J. (1993), *The theory of industrial organization*, MIT Press.

Van de Velde, D. (2003), "Regulation and competition in the European land transport industry: some recent evolutions", presented at the Thredbo 8 Conference, Rio de Janeiro.

Vickers and Yarrow (1988), *Privatisation: an economic analysis*, MIT Press.

White, P. (1997), "Experience in the UK bus and coach industry", in: Bradshaw, B. and H. Lawton Smith (eds.), *Privatization and Deregulation of Transport*, McMillan Press.

THE REGULATION OF TRANSPORT SERVICES AND INFRASTRUCTURE: THEORETICAL AND POLICY ISSUES

Marco PONTI
Polytechnic of Milan (IT)
Milan
Italy

TRANSPORT SERVICES: THE LIMITS OF (DE)REGULATION - ISBN 92-821-2345-6 - © ECMT, 2006

THE REGULATION OF TRANSPORT SERVICES AND INFRASTRUCTURE: THEORETICAL AND POLICY ISSUES

SUMMARY

Milan, February 2004

1. A FEW THEORETICAL ASSUMPTIONS

The traditional "social choice" approach states that public intervention is needed in the presence of social goals and/or of market failures. Historically, this intervention has assumed the form known as "command and control", via direct production or, more frequently, by means of public agencies. The generally poor performances of these agencies (generated by "capture", "rent-seeking" and "informative rents" mechanisms[1]) have, in fact, induced both the concept and the practical policy of public regulation. Command and control, regulation and market competition in turn can be seen within a subsidiarity[2] context: the former is to be employed whenever the latter fails to deliver.

A possible definition of regulation is the following: "*State intervention, aimed to reach welfare goals, by setting rules motivating efficiency-oriented actors*". This definition clearly implies that the State has special difficulties in combining welfare and efficiency objectives. Furthermore, "efficiency-oriented actors" may well be public enterprises, but this "orientation" is much more sharply focused in private (profit-motivated) firms.

The fact that the State faces problems in obtaining productive efficiency seems inherently quite natural: the minimisation of labour costs is an all-important factor of efficiency, while welfare objectives are in general oriented towards enhancing employment and labour conditions. But managerial skills are also compensated and motivated by profit more than by simple "good governance", which is the best possible outcome of public management.

Nevertheless, as we have seen, state intervention is needed not only in order to reach autonomous welfare goals, but also when the market fails to deliver productive or allocative efficiency.

Therefore, the very first issue is to define the proper scope of state intervention. And within the transport sector there indeed exists a wide range of situations where this intervention is needed: natural monopolies, externalities (both within the standard definition of externality, such as those related to the environment, and in the form of "club" externalities, like congestion), information asymmetries (related mainly to safety issues) and other special transport failures, like the Mohring effect or the existence of incomplete or inherently unstable markets. Income distribution can in some way be included in the scope of state intervention and, even if it cannot be defined as a market failure, it can be a legitimate public objective.

The question of which service has to be regulated (i.e. is in need of public intervention) and which one can be left open to competition, is theoretically straightforward: it depends on the political objectives and on the technical evaluation of the efficiency of the market.

The second issue is related to the choice between command and control public policies or regulatory interventions, within the definition proposed above.

As we have seen, within a classic social choice model, the public principal is assumed in fact to be both benevolent and all-knowing. Therefore, he will be perfectly able to obtain efficient results from his "agents" (public companies). Furthermore, his objectives will remain strictly and

unwaveringly aimed at welfare maximisation. But an assumption of public principals as being humans and not angels seems much more realistic[3].

Nevertheless, even if regulation (as against command and control) is assumed as the dominant strategy, it has to be kept in mind that its role is limited to a well-defined subset of public objectives. *Productive* efficiency is the main one, given that in this area the public principal faces severe conflicts of interests, as we have seen. A second set of objectives is related to natural monopolies and other market failures of the same kind (problems of efficient tariffs and access rules, etc.) which generate mainly *allocative* problems.

But it can be objected that other public goals cannot be kept at a strictly technical level (i.e. measured in terms of social surplus losses or gains), since their nature remains mainly political: strictly as with distributive and environmental issues, for example. Also in these cases, a regulatory attitude looks more effective than direct state intervention.

For example, if a country or region decides that local public transport has to be provided free of charge (while other services are deemed less socially relevant), this is a perfectly acceptable choice (but less so if these services are produced via command and control practices and not via competitive tendering).

Also, in the case of an opposing political choice, if a free-market provision of collective transport generates unstable results or dominant firms not justified by economies of scale, a proper regulation is again required, without any foreseeable need to return to command and control practices.

Environmental issues are, in theory, allocative failures (social surplus is not maximised, due to excessive consumption). Nevertheless, the same concept of externality implies a relevant distributive content (some actors damage other actors without due compensation). Furthermore, the uncertainties linked to the measurement of the related economic costs leave a wide area to political judgement.

But also in this case, the tools needed to reach environmental improvements have to be efficient, i.e. able to minimise the social costs involved in every environmental policy. And a regulatory approach looks by definition more efficient: vouchers and tariff techniques look far more promising than the traditional approach of imposing constraints and prohibitions.

In conclusion, while the area of public decision remains very large within the transport sector, the space for command and control practices (as an alternative to public regulation) seems to be shrinking, at least in theory.

2. THE SCOPE OF TRADITIONAL PLANNING IN THE TRANSPORT SECTOR

An important issue within transport policies which remains to be dealt with is planning instruments, even accepting the increasing role of regulation. The connections between land use, infrastructure planning and landscape control are the main areas where a more direct public role has to remain dominant. Low-density land use has been generated by mass motorisation via the increased accessibility of low-cost residential and commercial areas[4]. Low-density land use nevertheless makes

public transport provision very costly, public transport is generally subsidized, and more so where its full cost becomes unaffordable for many users.

Therefore, two external costs seem to be entrenched in low-density land use: public transport subsidies and the environmental costs of a more transport-intensive pattern of settlements (where individual transport becomes dominant). In theory, getting rid of any subsidy to public transport and at the same time internalising all the private transport externalities will solve the problem without any explicit planning activity. But this scenario is totally unrealistic, since this issue is also implicitly related with landscape values, which cannot be reasonably "priced"[5].

Regulation can well intervene here in optimising the construction and management process (public financing, concessionary regimes, etc.). In other words, regulation activity is called in to play its role in a later stage of the process.

3. THE SCOPE OF MARKET COMPETITION

Within the subsidiarity approach, which we have suggested at the beginning of this report, market competition has to be promoted until evidence of its failures emerges. Setting aside infrastructure operations, where only regulated or Demsetz[6] competition is possible, within transport services, the different modes offer a quite diverse picture.

Within the dominant land transport mode, road haulage is basically open to competition and no major problem exists, due to the very limited economies of scale and entry barriers which characterise this mode. The same pressure of competition and the social weakness of the operators (often small self-employers) generate problems of law-enforcement which must be improved: even stricter environmental and safety standards are possible but a re-regulation of this market is out of the question.

Remaining within the road mode, long- to medium-distance bus services are urgently in need of real liberalisation, at least in continental Europe. Long- to medium-distance buses compete with rail services for low-income demand, and these services do not have any real impact on the environment, nor any need of subsidies (to the contrary of rail and local services). Both users and taxpayers are severely penalised by this defence of (public) rail services. This is a very effective example of "non-benevolent princes", given the social characteristics of the patronage of these services. The situation, nevertheless, is now slowly improving.

Local public transport is quite a different case. Here, the British experience[7] seems illuminating. Full liberalisation has generated problems of unstable markets, followed by spatial monopolies (contestable more within the economics textbooks than in practical terms). The users have been penalised, as the quality of service has deteriorated. The theory supports these practical results: Mohring effects[8], network effects and other types of market failures are apparently working together with some characteristics of demand (related to information, the long-term effects of decisions on residential location and car ownership, etc.), in generating severe problems.

On the other hand, regulated (Demsetz) competition has delivered good results across the board (see the well-known London case). Moreover, since regulated competition, in terms of competitive tendering, can fully guarantee *any* social objective (even free transport, if so decided), the widespread European resistance to the opening of this type of market is another example of "non-benevolent princes" being "captured" by the interests of the suppliers of the service. In due time, even some form of full liberalisation may well be introduced, subsidizing the users[9] instead of the suppliers (this approach has not been tried in the UK) and carefully surveying the above-mentioned and ever possible undesirable consequences.

A far more uncertain picture comes from the rail sector. Here, even within the services, both economies of scale and sunk costs are present, together with the other problems (Mohring effects, etc.) mentioned above. Furthermore, rail services have strong interlinks with infrastructure operations, generating large transaction and severance costs.

There exists very little experience in the liberalisation of rail services. The British case has been quite particular in its form, and in any case not very successful, mainly due to serious mistakes in regulating the infrastructure (see also point 4.2.1). On top of this, there is little overall experience of the "free access" of rail services over a given track network (aside from a partial case in the USA).

The European liberalisation process has until now been reduced to limited entries within the freight sector over a time span of more than ten years. Nevertheless, the reason for this slow pace is far from technical in nature: liberalisation has been opposed with full success by the incumbent public companies, with the single states (their owners) protecting and helping this opposition.

The main problems here are twofold: the level at which separating services from infrastructure generates excessive transaction costs[10] and, as we have seen, the possible economies of scale, i.e. whether "natural monopoly" effects can arise even within the rail services.

The existence of the first problem is evident: for a subway line the separation of infrastructure from services has little economic sense. The rolling stock here is an essential asset, barely divisible from the infrastructure and lacking any secondary market. So, where does the threshold lie? Possibly in the presence of complex networks, where long-distance passenger services are operated together with freight and local services, separation is advisable. In the case of isolated lines with limited demand, separation seems a dubious choice, and a sound public regulation of a monopoly can well substitute for open-access strategies.

Economies of scale are different in nature: they are certainly present in rail services (rolling stock procurement in large quantities, maintenance, etc., are strong examples). But here any real experience of a free market is lacking, outside the USA, but even here with limits and specific constraints. In this case, a well-defined, dynamic policy can be suggested[11].

The situation looks similar within the air sector. Notwithstanding widespread declarations of liberalised markets, the sector is highly protected (and self-protecting). The slot regime is based on "grandfather rights", so that the most lucrative routes are plied only by incumbent companies, and the intercontinental services are in general not open to external competition. The operators naturally make use of cross-subsidies in these cases, and so the other markets are affected too[12].

The argument that large companies are suffering (even before September 11th) while new, low-cost entrants are prospering and therefore that competition is in fact at work, does not seem convincing at all. Large national companies have been suffering for many years from high costs, low

productivity and unsound fare policies; the only new development is that the states involved are certainly less ready now to subsidize them and in Europe there are growing constraints on doing so.

The low-cost companies are operating from minor airports and cannot take on the high-yield routes: i.e. these companies are growing *notwithstanding* the present barriers and, thanks to their inexpensive fares, are attracting low-income travellers (and more recently, budget-conscious business travellers).

A completely different structure of the entire air sector will probably emerge from a real liberalisation of this market. Little can be said of something which has never been experienced.

Even in this sector, economics of scale or of network may indeed play a relevant role. There are also some doubts about another type of market failure emerging, in the form of incomplete and therefore unstable markets[13], with the consequent need of some form of public regulation. But, first of all, a really competitive market has to be promoted, doing away with the "national champion" concept, which has nothing to do with efficiency and the protection of users. Only after this attempt, if problems of instability or incomplete markets emerge, can public regulatory intervention be properly considered.

Sea shipping may be a case of an unstable liberalised market, already operating for many years. The wide fluctuations in demand, supply and prices may certainly have generated some inefficient outcome here, but on balance the overall benefits of this competitive setting seem to make public intervention inadvisable, if only in order to protect the environment and, perhaps, the weaker labour components (as we have seen for the road haulage sector).

Table 1. **The subsidiarity chain in transport policy action**

Main areas		**Examples/Current issues**
Liberalization	• Transport services in general	• Long-distance rail and bus services • Intercontinental air services
Regulation	• Infrastructure operation/building • Unstable/non-existent service markets • Efficient charging and access rules	• Public–private partnership in infrastructure • Demsetz competition for local transport • Competitive tendering for concessions • Slot allocation
Planning (direct public intervention)	• Infrastructure design and location • Environmental and social values • Land use/transport policies	• European Common Transport Policy, TEN, etc. • Kyoto standards • Urban sprawl containment

4. THE SCOPE OF PUBLIC REGULATION

4.1. The issues

As we have seen, public regulation has to simulate the market pressures toward efficiency, where market competition cannot work properly. This is the case for natural monopolies, i.e. transport infrastructure. Club or co-operative solutions to this problem can only work in theory: transport infrastructures are in fact not only natural monopolies but also legal monopolies, in the sense that land use, of which they are a relevant building brick, is planned (under a command and control type of public intervention as we have seen above). But their operations and physical construction can be efficiently regulated, i.e. left to efficiency-oriented actors (basically private ones).

This is already the case for the pure construction activities, regulated by competitive tendering. Construction joined with operations, i.e. project financing practices, deserves a more in-depth analysis, as we will see later.

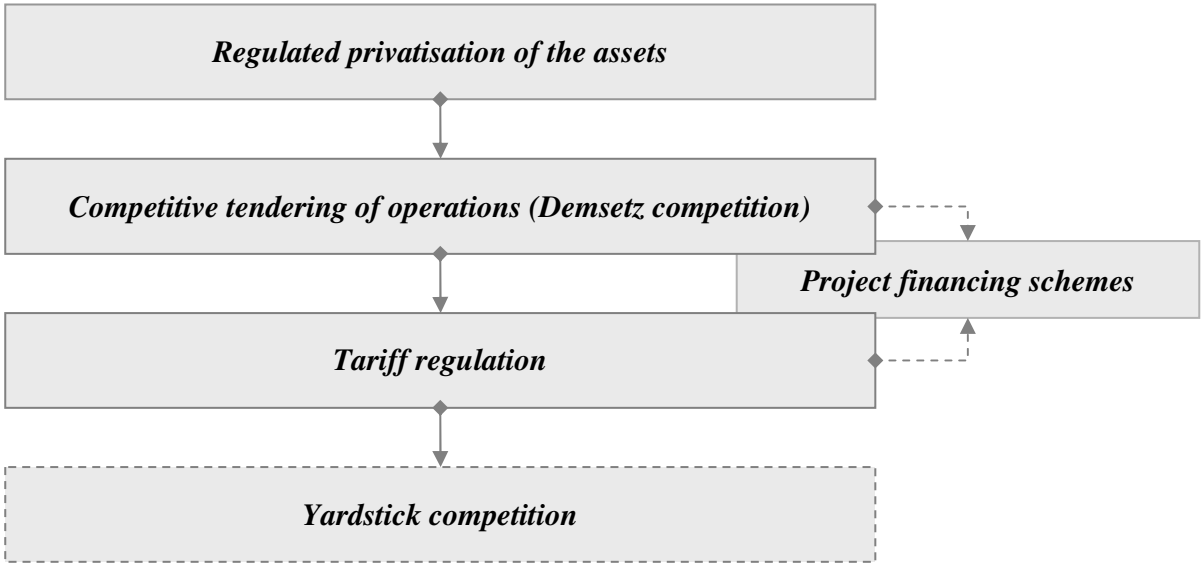
A proper regulatory regime for infrastructure is a highly complex task, with many aspects still to be tested and even fully understood. Furthermore, the resistance of the political actors[14] to pass from a command and control regime to regulation practices seems particularly strong (another proof, if still necessary, of the capture mechanisms so well defined within the already mentioned public choice approach).

4.2. The main regulatory policies for infrastructures

There exists a wide range of regulatory policies; the main ones are summarized here, in order of their degree of innovative content, i.e. in inverse order of their distance from the status quo. This can also be seen as a kind of subsidiarity chain.

This logical chain is somewhat different and more complex than the one proposed by Gomez-Ibanez in his recent book on the general subject of infrastructure regulation (2003), since private contracts, mentioned in that text as one of the main categories of regulation, are rare within the transport sector, while other issues looks far more relevant.

Table 2. **The subsidiarity chain in the regulation of transport infrastructure**



4.2.1 Privatisation of the assets

This is the radical British model for every public utility sector. The implicit risk for the public interest seems nevertheless quite high, given the option value implied in this choice, which is basically non-reversible. Capture risks remain paramount, given both the length of the public-private relationship involved (practically eternal) and the power held by a (generally) large private monopolist, so created by a public decision.

In railways, the UK experience has shown severe problems both in information control during the privatisation phase[15] (apparently, the real future costs of maintenance have been underestimated on purpose) and in the subsequent regulatory policy. The core issue is that a *private* natural monopoly is contestable as a property (others may buy it), but keeps too much power against its public regulator. In other words, this is a policy which again assumes a "benevolent, all-knowing prince" attitude. Periodic tenders for concessions look a far less demanding strategy, since the market pressure itself, and the transparency involved in the tendering process, help towards a more multi-faced control of the results.

For airport infrastructure (again, mainly a British experience) the problems seem less severe, even if developments in the long run remain on the table, in which this type of policy may face more complex problems (for example, varying land-use choices).

4.2.2 Competitive tendering of concessions (Demsetz competition)

We have already seen the advantages of this tool for transport services, when fully-fledged competition is deemed inadvisable. For infrastructure operations, the experience is still quite limited, but in theory it appears to be a balanced policy, limiting the risks of capture linked with very long public-private relationships. For some types of infrastructure, nevertheless, the length of the concession has to be fine-tuned, referring to the technical content of the assets involved, and the consequent need for sufficient learning time for the new-entrant company: for example, rail and air

infrastructures may well need longer concessions than toll highways (which have mainly a simple maintenance and toll collection content).

For infrastructure, it is quite obvious that keeping the same operator for a long time raises the risks of information asymmetries and capture phenomena.

Furthermore, long concessions for infrastructure are generally explained by the need for amortization of long-life investments. But this is a highly questionable argument for transport infrastructure: these assets (essentially civil works) have a practically *infinite* life, and therefore there is no physical amortization at play, only financial amortization (where applicable), and sound contractual constraints on maintenance standards and obligations seem to provide a sufficient controlling tool.

Therefore, the length of concessions can be limited, setting proper rules both for the incumbents and for the new entrants in case of change of concessionaire -- the possible consequence of a re-tendering process.

4.2.3 Building and operating concessions (project financing)

When a new investment is the main object of a concession, generally the practice in use sets a very lengthy concession period, assuming the need for a complete recovery of the invested capital. This practice has the well-known advantage of joining the responsibility for construction, operating and maintenance with the consequent overall optimisation of the entire system.

But we have already seen the risks of long concessions[16], and the weak rationale of linking the (assumed) physical with the financial amortization. Therefore this approach has to be considered with prudence, also given its capability of disguising public expenditure as private, via over-generous risk guarantees in favour of the private investors, which in fact transforms those investments into risk-free, sovereign loans. This was the case for the high-speed rail lines in Italy, but many other projects have similar contents, not easy to identify immediately, given also the ever-present possibility of reopening negotiations in the long term and far from a competitive context.

4.2.4 Tariff regulation

Tariff regulation is basically required in two cases: a) dealing with transport services when there are distributive, congestion or environmental issues involved; b) dealing with infrastructure, when productive efficiency has to be reached without competitive tendering (i.e. when the provider of the service is assumed to be unchangeable) and finally, in the extreme case of privatisation of the main assets. Price capping is the main technical tool used in these cases. Of course, mixed or overlapping situations are possible.

We deal below in more detail with the uncertainties related to tariff regulation, both within transport services (see point 4.3.) and for infrastructure (see point 4.3.3).

4.2.5 Yardstick competition

This strategy (also sometimes known as a "tournament") is certainly a form of simulated market, but looks by far the most conservative policy among the ones considered here, and remains quite close to command and control practices.

The regulator limits himself to comparing the results of different public companies in the same field (for example, airports or railways) and setting "prizes" and "punishments" according to their performances. So far, so good. The problem is that this approach is basically coincident with a sound command and control policy in the case when many operators are at play.

The problems related to insufficient incentives, the mixture of efficiency and welfare objectives, capture, etc., which gave rise to the regulatory evolution itself, remain fully present. Regulators and regulated subjects are not sufficiently separated and juxtaposed. Even for the Japanese railway reform (perhaps the largest example of a form of yardstick competition within the transportation sector), the model has been adjusted in order to guarantee a high level of autonomy for the different local companies[17] from the central regulator, with the explicit aim of minimising the risks of capture mechanisms[18].

4.3. Some technical examples of regulatory issues within the transport sector

4.3.1 Congestion charges and access rationing

Congestion implies a mismatch of demand and supply of transport infrastructure (access rationing is basically the same issue). Two main problems can be underlined here: the first is related to project financing.

The rationale of the construction costs of natural monopoly being charged to the users can be related with congestion charges; otherwise this charging generates a well-known welfare dead-weight loss. In turn, congestion charges are assumed to be, by definition, efficient and therefore the related revenues can efficiently (and equitably, see the club-externality problem[19]) be used for financing the infrastructure costs. But infrastructure suffers from indivisibilities, so in general it is underutilised at the beginning of its technical life and congested toward the end. Nevertheless, financial needs go the opposite way: they are maximal at the beginning and thereafter tend to decline.

This is another element which suggests maintaining a prudent attitude toward project-financing strategies: the traditional competitive tendering of construction contracts, followed by a sound periodic tendering of concessions for operations and maintenance, may often be a more prudent choice, where even the charges to the users can be kept under better control.

A second issue related to congestion is the (highly questionable) difference between the road mode and the controlled-access modes, i.e. railways, airports and ports.

Congestion on roads has to be regulated via social surplus maximising charges, which exclude the less-willing-to-pay traffic. It is assumed that, since congestion is non-existent (or minimal) within the controlled-access modes, for them no congestion charging is needed. This would be true, *if and only if* the excess demand in these modes is excluded by the traffic controller with a surplus-maximising rationale. But this is generally not the case: railway and airport (and ATC) capacity is controlled basically by grandfather rights, or similarly inefficient criteria.

Auctioning the capacity, or setting a rationing access tariff, are the only two possible surplus-maximising practices and are exactly coincident with a road-pricing approach. These two alternative practices differ only from a distributive point of view (the first one skims all the social surplus from the users in favour of the operator of the infrastructure, while the second one leaves part of it to the users).

4.3.2 The "minimal efficient dimension" issue

This is a kind of preliminary issue in regulating network infrastructures (in transport, toll highways and rail tracks, it can be considered a problem of horizontal unbundling, as compared with the vertical unbundling issue dominant in non-transport sectors). The issue arises because for these networks there is no market pressure to determine their efficient dimensions.

Furthermore, this issue is important because the efficient dimension also has to be minimal, in order to avoid the regulated holding excessive power over the regulator (again due to capture risks). So the issue at stake here is a problem of balancing the possible economies of scale against excessive power. (This excessive power in turn may well have a negative impact on the proper working of a Demsetz market of concessions, and not only on the regulator.)

Toll highway networks probably have very limited economies of scale, related only to the dimension of the maintenance centres. Therefore it is reasonable to split up the concessions into subsets of a few hundred kilometres each. (Toll collection tends to become highly informatised, and already several concessionaires operate smoothly in an automated way without any physical interruption in the collecting systems.)

But for the road system, the same concept of concessions as exists in the present experiences seems highly questionable. Concessions are now generally based on a set of toll links, or on a single link to be built and operated, etc. But the traffic structure within dense areas (i.e. in the European context) is mainly short-distance, and the demand for mobility is served by the entire local network, of which the toll links are just a component and not always the largest one in terms of capacity.

Within this picture, a toll level aimed only at cost recovery (investment, maintenance, etc.) or at *productive* efficiency at best, is far from optimal in terms of *allocative* efficiency. Congestion and environmental externalities determine an optimal allocation of traffic flows which is far from the one induced by cost-recovery tolls. And if one considers the possible economies of scale of maintenance and minor investments, an *area-based concession scheme* looks a much more sensible strategy.

Furthermore, an area-based concession may well include other critical components, for example, the management of traffic information for emergencies (as in the case of major accidents) and even ancillary activities like parking facilities and public transport prioritisation (streetlights, separate lanes, etc.). Also schemes for shifting the number of available road lanes from one direction to another in peak periods can become a component of a package of activities which perceives the road system of an area as an integrated service or utility.

These packages obviously have to be committed under competitive tendering, and the duration of the concession can be kept limited, in the order of less than ten years, thereby limiting the capture risks involved in longer concessions[20].

For rail networks, the picture is even more complicated, due to the fact that even experience of non-national infrastructure concessions is very limited. Nevertheless, it is certainly extremely unlikely that economies of scale will coincide exactly with national borders (thereby confirming the efficiency of the present dimensions of the majority of infrastructure concessions or direct state management[21]).

The Japanese experience tends to show that minimally efficient dimensions are probably nearer to a regional scale (for large countries at least), depending on the number of long-distance lines which have to be cut when separating the networks (generally few, compared to the local lines which remain within the same region).

Also in this case there is a long way to go, at least in Europe, where a strongly nationalistic rationale continues to predominate over economic efficiency.

4.3.3 Financial issues

The established rule of setting a proper rate of return for regulated companies is based in general on the calculation of the WACC (Weighted Average Cost of Capital) index. This index is needed to properly remunerate the invested capital, especially, but not only, when investments are financed through the tariffs in an explicit way, and not left within the price-cap mechanism (see the following point 4.3.5).

The correct evaluation of invested capital (or Regulatory Asset Base, RAB) within a concession regime is a highly controversial issue. In the first place, its magnitude has to be kept to a minimum: productive efficiency requires, for capital not less than for labour, that the resources employed are only the necessary and efficient ones.

But often a conflict of interests takes place within the public sector: in selling a concession, or in privatising an existing one, the State may be willing to maximise its revenues, and doing so may permit or even promote an RAB far larger than the minimum technically needed in order to operate the infrastructure efficiently; and this capital can be really of a limited amount if the physical assets are kept public, see point 4.2.1.

Furthermore, the actual price at which the concession is sold can be far higher that the book value of the capital required: its price may well represent the discounted value of future expected profits. In turn, if the sales value is in some way included within the RAB instead of the book value, there is a risk of a spiralling and self-induced increase in the values of the entire concession system, given that a normal level of profit on capital is guaranteed via the tariff mechanism. Also, this second over-evaluation problem may generate a conflict of interests within the public administration if short-term revenue maximisation prevails over efficiency and the protection of users from undue rents.

In turn, the definition of a proper WACC requires special attention: it is necessary to take into account the specific level of risk for every regulated sector. Within transport infrastructures, if the commercial risk is taken away from the concessionaire by the public regulator, the WACC has to be lowered in consequence.

Furthermore, it is advisable to define a target leverage level, in order to avoid the opportunistic composition of capital from the concessionaires. Finally, concessionaires which are floated (i.e. whose value is left to the judgement of the stock market) deserve special attention from the regulator, which is obliged to be extremely transparent and prudent in all its regulatory actions, in particular as far as the X parameter of the price-cap formula is concerned.

Also the inflation index within the price-cap formula has to be handled with care: there is a tendency to curb its level with regard to planned inflation, and to not adjusting it to the real one. But this is an improper tool for addressing efficiency: inflation is an exogenous factor for the regulated company and efficiency goals have to be addressed adjusting the X parameter, which holds this role by definition (see below).

4.3.4 Further price-cap problems: patterns and levels of efficient costs

The price-cap mechanism, although by far the best-known tariff regulation tool available for infrastructure concessions[22], faces several problems, of which a few are summarised here: the first one

is related to which type of risk has to be left to the regulated companies. In transport infrastructures, it seems reasonable to leave to them only (or almost only) the industrial risks and not the commercial ones (i.e. those related with the level of demand).

The basic rationale for this sectoral advice is linked to the exogenous nature of demand variations on transport infrastructures: these variations basically depend on the overall economic growth of the country and on national and regional transport policies (competing infrastructures and their tariffs, gasoline prices, liberalisation of services, etc.). In fact, if a company faces a risk which is outside its control, it has to behave on the safe side, maximising the relevant prices, etc. It is the same rationale which allows the regulated company a full recovery from inflation (within the price-cap formula).

A second problem is related to the efficiency parameter, which is in general included in the price-cap formula. Its definition requires an accurate benchmarking (even if efficient costs can only be known through a learning-by-doing process).

Within concessions of transport infrastructures, this is far from easy, giving the absolute dominance of monopolistic, inefficient examples from which the relevant data have to be derived. Even the speed at which efficiency has to be obtained (implicit in the X value) has to be estimated taking into account the specific constraints faced by each sector (labour contracts, etc.).

Obviously the starting base, set each five years (the regulatory lag) when the price cap is recalculated, is the *costs*[23] incurred at that moment by the concessionaire, and not its *revenues* (the objective of the mechanism is to make the users pay for efficient costs and only for them, allowing for a motivating factor which is linked to the possible extra profits gained in each five-year period by the concessionaire, and known as the regulatory lag, thanks to its efficiency). This periodic readjustment of the tariff is known as the claw-back procedure[24].

4.3.5 The regulation of investments

Price caps, or competitive tendering, in theory automatically guarantee the efficiency of the investments: only those capable of generating net profits will be implemented by the regulated company.

The problem here, as we have already seen, is that by far the largest part of the transport investments in infrastructure are *not* profitable in financial terms, and are generally decided by the public actors for a set of social objectives. As far as this decision remains outside the autonomy of the concessionaire, it is perfectly correct then to finance the investments with a public source of revenue. This source can be either direct transfers or an increase of tariffs on the whole network.

The first case is in general predominant for railways (and ports), while the second is in use for highways. Airport investments are in general self-financed, with some cases of state or local subsidy for the smaller ones.

But guaranteeing the funds for investments to a profit-oriented subject generates the well-known Averch-Johnson phenomenon[25], i.e. the pressure for maximising the level of guaranteed investments. In this way in fact, given a normal level of risk and a corresponding normal level of profits, the *total amount* of profits of the regulated company is also maximised. Therefore, special care has to be given to the evaluation of the social benefits of the (proposed) investments, to their design standard and to their costs, even if a competitive tendering is made mandatory (this is not always the case, and anyway is the subject of information asymmetries on the side of the regulator).

All things considered, *large* investments in the transport sector have to be kept basically within a command and control frame, especially if the benefits of project financing schemes are not fully guaranteed. This may well be the case for toll highways, which present a low technical complexity: as we have seen, the potential efficiency gains from integrated construction and operations seem limited, and a different, more sensible strategy can be recommended.

4.3.6 Safety and quality regulation

It is well known that a monopolist has to be regulated not only on the tariff (and access rules) side but also on the quality side, since there are no specific incentives towards quality (and sometimes safety) improvements, as generally in the case of market pressures.

The problem here is technically quite complex, to the contrary of price regulation which is, at least in conceptual terms, rather straightforward.

The quality and safety of transport services provided by transport infrastructures require not only specific experiences and benchmarking, but also a direct and active role for the users, who are the main stakeholders, and generally, those who pay for the services (the residual claimants).

There is up to now little experience on the involvement of these subjects, and there is a strong urgency to develop such experience. Setting abstract quality standards is useful anyway, but certainly not sufficient. Furthermore, the question remains open of objective measurements of safety and quality, in order to limit the potential costs of litigation, in case of disagreement between the regulator (and the stakeholders) and the regulated companies.

Finally, the proper balance between mandatory standards and well-designed incentives is another problem which deserves special consideration in quality regulation, while for safety, the standard obviously has to remain the dominant policy.

4.3.7 The problem of the number of tills

This issue is well known within airport regulation, but it is also present in the railways and highways infrastructures (the main difference is that a proper regulatory experience is practically absent in the latter cases).

The core of the problem can be summarised as follows: how complex should the regulatory action be? There are in fact trade-offs: a fine-tuned regulation may in theory be more efficient, but is less transparent and leaves less space to the regulated companies to develop general strategies of optimisation. Let us start with toll highway concessions. A double till is already present when investments are decided and financed on top of the regulation of tariffs.

If tariff regulation also takes into account congestion and environmental issues, we can speak of a triple till, i.e. three different tools of public intervention.

For airports, the dominant theoretical approach is known as double till: tariffs are price-capped on the *air* side (landing fees, etc.), while on the *land* side (commercial activities, parking, etc.), the possible monopolistic rents are skimmed via specific royalties, since is technically almost impossible to regulate every single price of the services on offer. (In the case of periodic competitive tendering, pre-set air-side tariffs reach the same "skimming" effect on the possible land-side rents.)

The single-till approach, used for London's airports, limits price-capping on the air side. But this generates a distorted price signal: since the price-cap periodically eliminates rents from the overall revenues, the tariffs on the air side tend to decline sharply as the rents on the land side rise. Therefore, the more traffic (i.e. congestion) an airport develops, the lower its air-side tariffs become, which is evidently inefficient in allocating the relevant traffic.

Within the rail sector, the problem lies between a double- or a triple-till approach[26]: the double-till approach suggests limiting state intervention to services (with subsidies for social or environmental goals) and to investments (again with subsidies). This approach implies that the infrastructure is not subsidized, i.e. its full costs are charged to the users (via track-use tariffs).

Since the infrastructure is a natural monopoly, with sharply increasing returns to scale, its full-cost charging (against the marginal cost suggested by the economic theory), generates welfare losses. (In turn, track-use tariffs have to be price-capped, or the concessions have to be tendered periodically, in order to obtain motivating effects on efficiency.) Obviously, the triple-till approach requires specific subsidies for infrastructure operations as well. Also, the subsidies to infrastructure operations have to be capped (with a specific subsidy cap) if the concession is not periodically tendered out.

The trade-offs involved here are particularly evident. Given the complexity of the sector, a triple-till approach risks to render opaque the public objectives entrenched in the sector. What is the final cost to the public purse of the entire system?

Furthermore, within the double till approach, the subsidies given to rail services may well include those that in the triple till case are earmarked for infrastructure operations. The only remaining advantage of the triple till is its effect on competition in rail services: given the high entry barriers existing in the sector, low track-use tariffs (i.e. priced at marginal costs) are definitely more pro-competition (without the need for explicit subsidies to the service operators -- not easy to muster within a competitive environment).

These examples can be extended to other infrastructure as well (ports, etc.), since the core issues are basically the same.

5. IMPACT OF PROPER REGULATION ON OVERALL TRANSPORT POLICY

5.1. Efficiency

As we have seen, the main goal of regulation is, almost by definition, an increase in efficiency. In the case of monopolistic public operators, these efficiency gains can be, as underlined in Chapter 1, both allocative and productive.

Within the mainstream theory, in the case of private, monopolistic operators, only allocative efficiency (and equity) is in play, since productive efficiency is supposedly motivated by the profit-maximisation objective of the private actor. But recent elaborations have demonstrated, quite

convincingly, that the implicit efforts[27] needed to obtain productive efficiency are, also for private operators, really maximized only within a properly regulated contest.

Lower costs and prices in turn permit, firstly, an easier social acceptance of other public actions: for example, in the road transport sector, efficient charges for congestion or for internalising external costs will be implemented with less opposition if other infrastructure costs are lowered by efficient operations. But, secondly, similar results are also valid for other, more explicit social objectives: for example, given limited public resources for local passenger transport, lower production costs induced by competitive tendering can definitely permit extended services or lower fares as an alternative.

5.2. Competition

As already described, regulation simulates market pressure on efficiency. But market pressure in turn also generates a powerful incentive for technical and managerial innovation.

The slow pace of railway innovation in Europe is strongly linked to insufficient incentives: public companies cannot go bankrupt, and wages and salaries are paid for centuries-old practices and technologies which are unable to follow the evolution of demand. The French railways are an exception, at least in terms of technical innovation but, setting aside the fact that the exception confirms the general rule, the public costs involved have been and probably still are extremely high, with shaky overall industrial results (exports, ailing rolling stock production, etc.).

In the case of "flagship" airline companies, the lack of any efficiency-oriented regulation has proved to be very unsuccessful (even if here the proper regulation is probably quite near to full liberalization). What has been the accumulated social opportunity cost of many years without low-cost alternatives (i.e. managerial innovation), which emerged only when a (still partial) liberalization was permitted?

And given the limited experience of inducing regulation in infrastructure, the potential for technical innovation in this sector probably still remains largely untapped (see also the outlines of innovative, area-oriented road concessions, proposed here under point 4.3.2).

Finally, proper regulation practices can reduce critical entry barriers in several sectors. Unbundling rail infrastructure from operations may be an effective example, since it substantially reduces the market power of the incumbent rail service companies. But similar situations are present within the port and airport systems as well, where traditional (even if less formal) alliances among dominant operators and infrastructures can be broken by a proper regulation of the latter.

5.3. Investments

In the previous sections, a position has been presented, with some elaboration, in favour of a rather traditional, planning-oriented attitude towards large investments, given the risks generated by the extreme distance of these economic activities from the conditions required for the efficient functioning of competitive markets.

Nevertheless, sound regulatory practices can also create benefits in this area. An example could be the reduction of overinvestment (gold plating) in airports. Too-expensive airports (and the examples are plenty) are an additional barrier to the entry of low-cost operators; the main hope here is that the reduced willingness to subsidize[28] large flag carriers (which actually pay for the top-grade airports) will generate spontaneous pressure to reduce airport costs across the board, at the same time

curbing unnecessary, luxury-oriented investments. But similar situations are equally present for some railway infrastructures, where less than essential investments abound.

6. A FEW FIRST-HAND ITALIAN CASES OF REGULATORY START-UPS

6.1. The administrative context

The creation of an independent regulatory authority for the transport sector was proposed at the beginning of the nineties, within the same parliamentary document which proposed the creation of the electricity and telecommunications authorities (which were, in fact, created a few months later in line with true privatisation of important segments of these industries). But the transport sector appeared highly fragmented, and for only a minor part of it was a privatisation process under way -- the larger highway concession -- and even this sale was limited to operations and some investments, excluding the ownership of the main assets.

The proposed authority was therefore shelved and replaced by a consulting body, internal to (and responding only to) the Ministry of the Economy, with very limited resources and a few external experts (NARS)[29]. Its technical role seems ill-defined: it mainly deals with infrastructure (but excluding ports) and some services are included.

This weak body was seen by the existing technical ministries in charge of the different utilities as a useless obstacle to their consolidated administrative role (setting tariffs, giving concessions, etc.). Therefore, the real counterparts of NARS were two-sided: the regulated companies, generally in close alliance with the sectoral administrations. The existence of capture phenomena, it has to be remembered, is by definition the main reason why regulation is supposed to substitute for command and control practices.

Local public transport was not included: since the sector was, incorrectly as we will see, supposed to be on the verge of a substantial liberalization process (even if in terms of Demsetz competition), its regulation was left to the anti-trust Authority which, by Italian law, is not allowed to regulate matters where market competition cannot take place (either for technical reasons or for political decisions).

6.2. Railways

An attempt was made to raise a debate on the proper size (see the issue of the minimum efficient dimension) of governance and the length of concession of the national rail infrastructure, but a strong political veto stopped any further elaboration on this issue[30].

Therefore, regulatory activity for the infrastructure has mainly concerned tariffs for the allocation of track capacity. The methodology is of the transfer-cap type, since transfers are the main source of revenue covering rail infrastructure operation costs (about two-thirds); furthermore, there is a constraint on allowing profits.

The implicit economic rationale of a tariff of this type lies somewhere between *marginal* and *average* cost pricing, which is a compromise with some merits given the problems arising with both of the extreme choices. In fact, short-term marginal cost pricing for rail infrastructure, generating very low tariffs, implies an extremely high subsidy but clearly lowers an important barrier for new-entrant competitors.

The contrary is obviously true for average cost pricing (i.e. high tariffs), with the additional drawback of incurring deadweight losses (loss of social surplus) and some environment-related problems, due to the diversion of rail demand in favour of other, less environment-friendly modes.

The system is now fully operating, even if its initial, main objective of setting efficient rules for competing operators, both for freight and passengers, has been contradicted by the shelving of any liberalization for passenger services and the heavy constraints remaining on competition for freight services -- mainly linked to the control of the incumbent both on national services and on the main terminals.

Nevertheless, the issue of scarcity tariffs (congestion is not the more proper term for regulated access modes) has not yet been addressed.

Long-distance passenger tariffs are also regulated by NARS, within a price cap system; the main problem here is that the rules set by NARS have, in fact, been drained of practical effects by an ill-conceived goal of curbing inflation (rail tariffs have been frozen for two years). In fact, the impact of rail tariffs on inflation is minimal, so the real objective of this action seems to be more related to short-term political consent (crowd-pleasing), than to overall economic considerations.

A specific issue which is also not included within the present regulatory action for railways concerns cross-subsidies between efficient and inefficient passenger services. Officially, since no competition takes place for these services, no third party is damaged (excluding perhaps the users of efficient services, who subsidize the users of inefficient ones): the standard rationale for public intervention against cross-subsidies implemented by dominant operators is their use as a barrier against the entry of competitors, which in this case is not allowed anyway.

Why does the monopolistic, public operator of these services keep the loss-making ones, without having any formal constraint to do so? And why can this be criticized from a public-interest point of view?

The answer to the first question may well lie within a capture phenomenon (related to informative asymmetry). Loss-making services are probably cross-funded in order to keep an implicit pressure on local and central politicians to indefinitely postpone any liberalization process, hinting that the loss-making services will no longer be possible within a competitive context.

In fact, if there is a political will to subsidize long-distance rail services for social reasons, this can be achieved, even in a system of competitive tendering, by setting explicit subsidies for these services. But then the possibility of obtaining the same social objectives via less-expensive solutions will emerge (buses for shorter distances and subsidized air fares for longer ones). The public interest is damaged twofold: firstly by the loss of democratic transparency in the allocation of subsidies and secondly by the (very likely) inefficiency of long-distance rail services with reduced demand.

6.3. Airports

At the start of some privatisation of airport concessions (Rome, Turin, Naples), the issue of concession length and assignment procedure was raised, but a strong political opposition emerged, both against any opening up of competitive tendering and against any hypothesis of concessions being limited in time.

The reasons given are the need to defend the sector from foreign attacks, and the priority of promoting investments, assuming (incorrectly, see point 4.2.2) that the amortization period has to coincide anyway with the duration of the concession concerned. Therefore, a general rule of forty-year-long concessions, without any competition, has been set by law; a situation which recently even raised explicit criticism from the Anti-trust Authority, which is not officially entitled to intervene on regulated (natural in this case) monopolies.

As with railways, even for airports the regulatory issue has been limited to tariffs. A single till price cap mechanism has been developed (on the air side only). This type of regulation is obviously much more favourable for the regulated company than the double till scheme, since it leaves uncapped the potential extra profits generated on the land side (commercial activities and parking).

Here emerged in full light the weakness of a regulatory body devoid of any sanctioning power, like NARS. The price-capping mechanism requires, as a minimum, a budgeting and accounting procedure by the regulated companies which is suitable for measuring the relevant factors involved (mainly industrial and financial efficiency levels, etc.). The format for these accounting procedures was given to the companies and the price cap mechanism was approved by the appointed Ministry of the Economy. But nothing happened thereafter, since the proper accounting figures were never delivered. The resistance of the airport lobby has apparently been able to block any attempt to lawfully enforce the price-capping procedure (and the corresponding accounting duties).

Furthermore, the privatisation process is continuing, and this appears to be a worrying pattern indeed. In fact, the value of a concession depends directly on the expected streams of future revenues, which in turn directly depend on the regulatory regime.

In a situation of uncertainty, the same fact that concessions (or shares of them) are sold more or less referring to their present profitability levels, implies that both the sellers and buyers are confident that a real regulatory reform will never take place (or, worse still, that it will be possible to curb it enough). Here, a conflict of interests is especially evident: the present owners of airport concessions, mainly local administrations, want to maximise their revenues, showing no interest whatsoever in protecting users from undue rents (but this attitude seems really short-sighted, since it is in fact reducing the possibility of attracting more air services, particularly of the low-cost type).

6.4. Toll highways

Italy has a long tradition, dating from the fifties, of concession of toll highways, built and operated by public agencies. Now the system has more than 6 000 kilometres of extensions, operated by a dozen concessionaires (as with airports, not owning the physical infrastructure).

The regulatory issue became especially relevant towards the end of the last century, when EU rules made mandatory the privatisation of the main concession (Autostrade SpA), since it was part of a heavily subsidized public industrial conglomerate, which had to be dismantled. Autostrade had about 50 per cent of the network, and far more than that as a share of traffic.

The privatisation process started immediately under the above-mentioned conflict of interests: no public debate nor economic evaluation of efficient dimensions, and the concession period extended for forty years without any competitive tendering.

A regulatory contract has been defined, both for operations and investments, formally based on a price-cap mechanism. Nevertheless, large investments were explicitly included, to be paid for by increases of tariffs on the entire network, and therefore not self-financing with their own revenues. This is a first contradiction of the price-cap approach, which implicitly assumes that only "efficient" investments will be induced. Other investments have to refer to different regulatory mechanisms, never made explicit in the contract.

A second, more severe, contradiction in the contract guarantees the buyer of the concession a balanced budget, including an ill-defined rate of return (not even based on WACC). A price cap may well generate temporary extra profits but, in the case of an inefficient operator, may generate proportional temporary losses.

The sale of the concession obviously produced a large sum, partly as book capital (real assets like buildings, machinery, etc.) and partly as the discounted value of expected profits.

In the first regulatory lag, the concessionaire made extremely high profits (it was rated as "the second company in Europe in terms of rentability" by one of the main financial commentators in Italy[31]).

At the end of the first regulatory lag, the above-mentioned contradictions came to full light, and a severe conflict arose between the Ministry of Infrastructure (technically responsible for the concessions) and NARS. The Ministry strongly defended an interpretation of the contract totally in favour of the concessionaire, while the secretary of NARS objected.

Setting aside a long list of other technicalities, the core of the conflict is referred to as the claw-back mechanism (see point 4.3.4): the concessionaire stated that since it was not made explicit in the contract but only a price-cap system was mentioned, this was implying that no claw back was supposed. But it is evident that a price cap unable to limit in time potential extra profits is not defending the users from monopolistic, undue rents.

This conflict was in fact extended to include the positions of two different parties of the governing coalition, making evident the damages implied in the absence of an independent regulatory authority.

Political negotiations lasted more than one year, with a very poor result. The position of the concessionaire turned out to be the winning one, but with a very peculiar twist: the tariff increase was postponed for six months, due to the pressure of consumers' organisations (and due probably to the imminence of local and European elections). This created a loss to the concessionaire estimated at 22 million euros, which at this point, given the fact that its interpretation of the contract is now assumed to be correct (even if only at the political level), remains totally arbitrary.

The overall political message which emerged from this process was that the defence of consumers is in no way left to a set of regulatory rules, but returns, if so chosen, into the hands of the political sphere. (Anyway, the undue rents calculated by NARS and left to Autostrade are larger by *an order of magnitude* than the savings for the users resulting from the above-mentioned delay.)

The damage here is much larger than in the case of airports: in fact, not only within the highway sector but for all the regulated monopolies, the new rule emerging is the absence of rules: direct negotiations and lobbying power will be the name of the game. But this may well also have a negative impact on the overall privatisation process: the absence of rules can be a source of large profits, but at the same time the level of political risk becomes very high.

6.5. Local public transport

In Europe, after the UK, Italy has had a rather innovative attitude in this sector, at least formally. A specific law of 1997 set out a general rule of competitive tendering for local transport concessions. Two immediate problems emerged: a mandatory date was not clearly defined and at the same time the financing of the entire sector was transferred from central administration to regional level (together with the same amount of funds, albeit not earmarked for the transport sector).

The reformers who defined the new administrative and financial setting[32] were hoping that the lack of earmarking of the transferred funds would generate an opportunity cost at regional level (related to the possibility of alternative uses). But not a single region opened a debate on the level of subsidy allocated to the sector, nor on its production costs, definitely higher than in other European countries (particularly in terms of labour costs at PPP). The level of subsidy averaged, and still averages, 65 per cent of the revenues.

This lack of debate (why 65 per cent in every region, with only slightly higher values in the south of the country and lower values in the north? -- why not 80 per cent or 40 per cent?) was a clear indication of widespread capture phenomena, signalling that the real public objectives were not related to the demand side of the service (income distribution, environment, etc.) but to the supply side (employment, salaries and wages and, in turn, votes). This has been largely confirmed by the following developments.

In the first place, the regions delegated the competitive tendering process to the municipal level, in this way maximising the conflict of interests, since the municipalities are generally the owners of the existing concessions (i.e. are the incumbents). Secondly, a long series of postponements followed. Furthermore, the regions set specific rules protecting the workers in the sector as would be unimaginable in the private sector: the new entrant was compelled to employ all the workers of the incumbent, with the same wages and work regulations (productivity is generally low in the sector). In this way, the incentive for competition practically disappeared (especially if one also takes into account the "information rents" of the incumbents).

Few tenders have been implemented up to now, and by far the winners have been the incumbents (in many cases, under the fragile disguise of a different name, generated by the nominal participation of some external or even foreign company).

Eventually, the new government entirely abolished the constraint of competitive tendering for concessions (which remains just a possible option) and further postponed the entire process.

The more evident capture signal can also be referred to social considerations: Demsetz competition was the only option open for the Italian case, and this type of competition allows for the complete protection of any type of social choice (service network, frequency, stops, tariffs, etc.). If public funds are limited, as they actually always are, less competition means higher production costs and therefore either reduced services or higher tariffs.

7. CONCLUDING REMARKS

Public regulation of transport services and infrastructure is a highly complex task, and basically still in its infancy. Command and control practices dominate, even when they are no longer needed. The liberalisation process, in turn, is slowed down by extended capture phenomena.

A first point has to be underlined: regulated (Demsetz) competition does not conflict with social objectives. Even free transport can (and must) be provided within a competitive context.

The main tenet of the problem is the following: direct intervention (command and control), regulation and market competition have to be considered within a subsidiary approach. Why does this approach make explicit a definite hierarchy of strategies? The traditional assumption, known as "social choice", of a benevolent, all-knowing prince is no longer acceptable, even if the perfectly egoistic prince contained in the public choice scenario is also too extreme. A balanced attitude has to stay on the safe side: if possible, don't assume that the prince is necessarily benevolent or fully informed.

Nevertheless, public regulation and market competition are not so far apart as is commonly considered. The market is not the absence of rules and constraints, quite the contrary: it has been built as a complex set of rules and laws, which have taken a couple of centuries to be properly established and which are under continuous evolution as social values (and the technology of the sector) evolve[33].

There is a large ideological difference between liberal values and pure *laissez-faire*. Furthermore, public regulation itself is not a purely technical issue: in reality, implicit in regulation choices are different visions of economic democracy and of social priorities.

The transport sector is quite peculiar in this sense: it is full of market failures and involves very critical values and social objectives (freedom of movement, the environment, safety and security, etc.).

The stronger the drive to liberalisation, the more (the necessary) public intervention has to be attentive and up-to-date; in other words, the more "market" we want, the better "state" we need. Nothing is really spontaneous in market competition: it is a political construction, and much work remains to be done within the transport sector.

NOTES

1. See Buchanan (1969).

2. A term largely used within the European Commission policy papers.

3. This has to be noted, not only within the radical context of a public choice setting, where the public principal is presented as a standard *homo oeconomicus*, maximising egoistic objectives. Even within a more relaxed setting, where the mix of egoistic and altruistic objectives may be varied and, *ex-ante*, basically unknown, a prudent attitude tends toward some scepticism in assuming a pure "benevolent, all-knowing prince" hypothesis.

4. See Litman (2002) and Maffii, Ponti (2002).

5. Imagine pricing a Tuscany renaissance landscape, menaced by a dozen high-rise condominiums… . Infrastructure planning has similar problems: on top of the all-important landscape issues, here the natural and legal monopoly phenomena are also present, as are regional development objectives. Moreover, for setting priorities and therefore supporting planners in taking into account efficiency objectives, the traditional cost-benefit analysis provides important clues.

6. See Demsetz (1968).

7. See Banister (1997) and Fawkner (1999).

8. See Ponti (1997).

9. See Ponti, (2002a and 2002b).

10. See Gómez-Ibánez (2003).

11. Whether economies of scale play a dominant role is up to the market to decide. So public regulation has to be focused on breaking all the possible entry barriers (technical, financial, informative, etc.), even helping the implementation of a secondary market for rolling stock (see the British ROSCOs). If a dominant company emerges, thanks to its long-range lower costs, so much the better for the users; the regulator has only to avoid abuses of dominant position (i.e. the setting-up of a "Microsoft-on-wheels"). Given the actual role of the dominant, inefficient public companies, there is surely a long way to go before a dominant rail company, based purely on its competitive merits, will emerge.

12. See Doganis (2001).

13. See Tucci (2002).

14. See Ponti (2001).

15. See Nuti (1997) and CESIT (1998).

16. See Ponti, Federtrasporto (1996).

17. And in fact the regionalisation process itself can be seen as a form of yardstick competition, where, even within a command and control structure, every region becomes a residual claimant for the resources involved, and may well compare the results of other regions. This is what happened in Germany with the decentralisation of local rail services, where the DB national rail company had to face pressure from different, budget-minded regions and had to provide more efficient services; this decentralisation, in the end, has even set in motion a real competition mechanism, with the rise of new entrants, both public and private.

18. See *Japan Railway and Transport Review*, 1994.

19. A club externality occurs when the damaged subjects belong to the same social group as the damaging ones.

20. See Newbery, D.M. (1998), *Fair and Efficient Pricing and the Finance of Roads,* University of Cambridge.

21. See Preston and Root (1999).

22. See Marzi, Prosperetti and Putzu (2001).

23. WACC included, as normal profit.

24. Strange as it may seem, this obvious statement in important cases (see, for example, the Italian highway infrastructures regulation, point 6.4.), is not fully accepted, with large and undue extra profits for the concessionaires, which thus prove themselves perfectly able to capture the regulator (also thanks to the far from minimal dimensions of the concessionaires).

25. See Averch-Johnson (1962).

26. See ECMT Round Table – Thompson (2002).

27. Efforts are, in fact, a form of cost, even in practical terms.

28. As we have seen in point 3.

29. *Nucleo per l'Attività di Regolazione dei Servizi sociali*, for which the author has been an external expert on transport (highways, airports, railways and maritime services) for three and a half years, under two different governments.

30. It was an attempt made personally by the author in his role of advisor to the Minister in charge (Mr. Bersani).

31. Penati, *Corriere della sera*.

32. The author collaborated in it to some extent, being at that time economic advisor to the Transport Minister.

33. A good example comes from the former Soviet Union: the dissolution of the State has generated a highly distorted economic structure (organised crime, etc.). In that country, the task of reconstructing proper market rules looks much more challenging than the destructive phase.

BIBLIOGRAPHY

Averch, H. and L. Johnson (1962), Behaviour of the Firm under Regulatory Constraint, in: *American Economic Review*, Vol. 52.

Banister, D. (1997), Bus Deregulation in the UK, in: McConville, J. (ed.), *Transport Regulation Matters*, Pinter, London.

Buchanan, J.M. (1969), *Cost and Choice: An Enquiry in Economic Theory,* Markham, Chicago.

CESIT (1998), *Liberalizzazione e organizzazione del trasporto ferroviario in Europa*, Rapporto di sintesi, Rome.

Demsetz, H. (1968), Why regulate utilities?, in: *Journal of Law and Economics*, Vol. 11.

Doganis, R. (2001), *The Airline Business in the 21st Century*, Routledge, London and New York.

Estache, A., Ginés de Rus (2000), *Privatization and Regulation of Transport Infrastructure.*

Fawkner, J. (1999), *Buses in Great Britain, Privatisation, Deregulation and Competition*, London Transport.

Gómez-Ibánez, J. (2003), *Regulating Infrastructure. Monopoly, Contracts and Discretion,* Harvard University Press.

(1994), Domestic Transport in Japan Present and Future, *Japan Railway and Transport Review.*

(1994), Restructuring Railways, *Japan Railway and Transport Review.*

Kerf, M. (1998), *Concessions for infrastructure – A Guide to Their Design and Award.*

Litman, T. (2002), *Evaluating transportation land use impacts*, VTPI, Victoria.

Maffii, S., M. Ponti (2002), *Pianificazione dei trasporti e del territorio: effetti attesi ed effetti perversi*, Ricerchetrasporti, Milan.

Marzi, G., L. Prosperetti, E. Putzu (2001), *La regolazione dei servizi infrastrutturali*, Il Mulino, Bologna.

Nuti, F. (1997), Il caso britannico, in: *Nomisma, Liberalizzazione e privatizzazione nelle ferrovie europee*, Vallecchi, Firenze.

Ponti, M. (1997), Le esternalità di consumo nei trasporti collettivi, *Economia e Politica Industriale*, No. 96, Franco Angeli.

Ponti, M. (2001), The European transport policy in a "public choice" perspective, in: *Proceedings of 9th World Conference on Transport Research*, Seoul.

Ponti, M. (2002a), Tpl, i sussidi, al 70% dei costi non sono una scelta scontata, in: *Trasporti,* supplemento quindicinale de *Il Sole 24 Ore*, No. 19.

Ponti, M. (2002b), Per sostenere il TPL il sussidio non è una strada obbligata, in: *Trasporti*, supplemento quindicinale de *Il Sole 24 Ore*, No. 21.

Ponti, M., A. Gervasoni (1996), Il finanziamento delle infrastrutture, in: *Ricerca Economica e Trasporto*, Federtrasporto, Centro Studi.

Preston, J., A. Root, D. Van de Velde (eds.) (1999), *Railway Reform and the Role of Competition: The Experience of Six Countries,* Ashgate.

Thompson, L. (2001), Railways in Eastern Europe, in: *ECMT Round Table 120, What Role for the Railways in Eastern Europe?*, OECD/ECMT, Paris.

Tucci, G. (2001), Ricerca di stabilità dell'equilibrio di lungo periodo nell'industria del trasporto aereo internazionale, 42° Corso Internazionale ISTIEE, *Nuova regolazione dei trasporti: principi e strumenti*, Università di Trieste.

SUMMARY OF DISCUSSIONS

Andreas KOPP
Chief Economist
OECD/ECMT JTRC

SUMMARY

1. INTRODUCTION

Since the end of the 1970s, strong political forces have led to a reduction in public sector production in the vast majority of countries around the globe. In the US, for example, the share of production in fully regulated industries decreased from 17 per cent in 1977 to about 6.6 per cent in 1988. Besides banking and insurance, telecommunications, public utilities and oil and gas extraction, it was the transport sector which underwent substantial regulatory reform. Significant reform steps were implemented for airlines, railroads and trucking, but only partially for pipelines, urban transit and inland waterways.

The motivations for the deregulation movement were strong and varied. Beyond a rather general expectation that deregulation and the substitution of public by private production would lead to efficiency gains, fiscal arguments have played a dominant role, in particular for transport infrastructure investment. Deregulation and privatisation were supposed to lead to a higher level of investment and a reduction in demands for public funds. For some parts of the transport sector, and in particular with respect to the provision of infrastructure services, deregulation is still evolving in many countries. For others, however, there seems to be increasing political support to re-regulate. Where systematic empirical information is available, it suggests that the fiscal objectives in particular have not been achieved. The volume of private capital that has been attracted to transport infrastructure has remained low relative to overall transport infrastructure investment. Moreover, the contractual relations between governments and private providers of infrastructure services seem to have been far from stable. Overall, deregulation and privatisation seem to have had limited success in reducing the need for allocating public funds to the finance of transport infrastructure.

The first part of the Round Table discussion was devoted to clarifying the form and degree of regulation required for the transport sector and for the provision of infrastructure services in particular. It centred on the question of to what extent genuine competition is to be expected in the transport sector and to what extent the decreasing average costs of transport firms form the basis of natural monopoly powers. This part of the discussion will be reported in section 2. In section 3, we set out the second part of the discussion, focusing on recorded efforts towards deregulation. Besides their limited success in achieving their fiscal objectives, regulators seem to hold a relatively weak position *vis à vis* the regulated firms: renegotiation of regulation contracts seems to be frequent and the bargaining power of the regulated firms appears to be increasing over time.

The Round Table concluded with a discussion of new concepts of regulation to strengthen the position of the regulators in pursuing the interests of transport system users, as well as those of present and future taxpayers. The last section sets out some general conclusions drawn from the discussions.

2. THE LIMITS OF COMPETITION IN THE MARKET AND FOR THE MARKET

Historically, the reason for transport infrastructure services provision being associated with the public sector domain has lain in the view that many services are "natural monopolies". The indivisibility of transport infrastructure facilities in general implies decreasing average costs, with an increase in the number of users. In most geographical settings it would therefore be technically inefficient to have more than one local supplier of transport infrastructure services. It was believed that a public monopolist was much more likely than a private one to pass the cost advantages of a single supplier on to the consumers by charging prices that reflected costs. Given these beliefs, it was natural to give the supply responsibilities to state-owned monopolies. The Round Table discussed the problematic features of the transport sector, where infrastructure policies were based on what might be seen as an over-generalised natural monopoly hypothesis for the sector.

The historical "natural monopoly industries" were and are typically composed of both potentially competitive segments (for example, railroad rolling stock), where competition in the market may be an effective way to allocate resources, and other natural monopoly segments (for example, railroad track) where competition in the market would lead to a high degree of monopoly power. Vertical integration has expanded monopoly from one horizontal level (the natural monopoly level) to other, potentially competitive segments.

As has been argued convincingly by Knieps (2005), it is crucial for successful deregulation to single out the potentially competitive from the natural monopoly segments. Strict criteria have to be applied to single out the latter. Following the terminology of the US antitrust legislation, they must constitute "essential facilities", or a monopolistic bottleneck. Facilities are considered to enjoy network-specific market power if:

- they are indispensable for reaching customers and/or competitors to access complementary markets;
- there is no substitute for using just one particular facility;
- costs are high relative to the market potential such that an equilibrium with additional providers does not exist.

2.1. Market entry and contestability

Even if these technical characteristics are given, potential entry might obviate the need for regulatory intervention. In fact, part of the deregulation policies were based on the expectation that markets for transport operations and/or the provision of infrastructure services would be contestable. (On the concept of contestable markets, see Baumol, 1982a and Baumol, Panzar and Willig, 1982b).

Contestable markets are those where suppliers do not enjoy monopolistic market power, even if their number is small, due to the threat of market entry by potential competitors. To avoid market entry, incumbents supply under conditions which are similar to actual competition in the market, and efficiency is thereby ensured. However, crucial conditions for the contestability of markets do not necessarily hold for the transport sector, in particular for the provision of infrastructure services (Stiglitz, 1987).

– First of all, at least part of the infrastructure investment costs are sunk. As infrastructure facilities are immobile and have no or imperfect resale markets, suppliers cannot recover the invested resources by selling the facilities in case they decide to exit the market.

– The existence of monopoly rents of incumbents may not attract entry. Potential entrants will anticipate what happens after entry, i.e. the fact that the positive profits of an incumbent monopolist may entail negative profits for an entrant who attempts to contest the market. In other words, a post-entry equilibrium might not exist due to the indivisibilities of the facilities[1]. Potential competition is then no substitute for actual competition.

– With sunk investment costs, entry may not lead to competition but to tacit or explicit collusion, driven by each firm's understanding of what is in its own interest.

– Attempts to avoid potential competition might lead to an inefficient choice of technology: technology and capacity decisions of incumbent providers will be made, aiming at the pre-emption of further entry (cf. Davidson and Deneckere, 1986 and Kreps and Scheinkman, 1983). Potential entrants, on the other hand, will try to make technology choices that facilitate collusion, which may differ from least-cost technologies.

Neglect of these problems has led to support for simplistic privatisation and "complete" deregulation policies which have proved to be unrealistic or ineffective policy options, at least in transport infrastructure sectors (Joskow, 1999).

2.2. Vertical disintegration

The identification of the sector segments with potential competition and the isolation of the genuinely monopolistic sections has often been the reason for jumping to the conclusion that these parts should be vertically disintegrated. The most prominent discussion of whether such a conclusion is justified concerns the separation of railway operations from the provision of railtrack services.

While a general analysis of the transaction costs of potential trade relations between operators and infrastructure service providers is still unavailable, it seems that deregulation experience has led to less support for vertical disintegration, in particular in the rail sector. The analysis which led to the proposal for vertical separation has, as is now emphasized by the sceptics, neglected transaction costs which are caused by trading transport infrastructure services on markets. In fact, in view of the widespread advocacy of the separation of infrastructure services from operations, its actual role in railway policies has been limited. It is an important part of EC transport policy, but separation including privatisation of the provision of infrastructure services has only been implemented in Britain. Interestingly, this form of rail sector restructuring is unique to Europe (Nash and Toner, 1999). Given the strong theoretical support for vertical disintegration, this is all the more surprising, as only in the US can we speak of competition in the market between vertically integrated railway companies.

On the down side of vertical integration and savings in transaction costs is the increased cost of ensuring competition in the potentially competitive market segment. Even with vertical separation of transport operations and infrastructure services, one of the most difficult, contentious and important tasks which regulators must confront is defining the terms and conditions under which operators should have access to the regulated "bottleneck" infrastructure facilities they require to serve their customers.

If transport operations and provision of infrastructure services are vertically integrated, regulators are confronted with enormous information costs to define the terms of non-discriminatory access to infrastructure facilities. A basic pre-condition for establishing non-discriminatory access is the existence of transparent cost accounting systems with a clear separation of variable and fixed costs for the different integrated activities. The cost accounting framework has to allow the detection of cross-subsidization and the verification that infrastructure pricing is not discriminatory against competing operators.

Access to infrastructure has, however, more dimensions than just pricing. The quality of equipment provided, the information on demand and customer characteristics, the co-ordination of timetables for railways, etc., all offer opportunities for discrimination. Non-price terms and conditions of access to infrastructure often involve subtle issues which are even more difficult to identify than price discrimination.

2.3. Competition for the market

Even if natural monopoly elements remain in the transport sector, competition can still be theoretically exploited by putting up for competitive bids the exclusive right to provide the service with natural monopoly characteristics, offering medium to long-term franchise or concession contracts. Based on the seminal work of Demsetz (1968), some advocates of deregulation and privatisation argued that such competitive bidding would lead to efficient supply, obviating the need for regulation. While franchise contracts have been successful for some areas of public service, even with relatively small transaction costs to set up the contractual relation and a relatively minor scale of investment, they have not been a panacea for transport infrastructure projects (Ponti, 2005).

First, the durability of infrastructure investment requires long-term franchise contracts to recover capital costs. In many countries, the state of financial markets, i.e. the unavailability of long-term credit, is a serious obstacle to setting up long-term franchise concessions. Short-term finance for long-term infrastructure investment confronts private investors with the double risk of dramatically changing capital market conditions, plus the risk that the government might not maintain its commitment to the public-private partnership. In both cases, investors may be unable to recover the high share of sunk investments in transport infrastructure.

On the other hand, the high costs of terminating a franchise contract for the provision of transport infrastructure services, and having a new competitive bidding process, limit the possibilities of a contract administrator credibly commiting to a specific, detailed, *ex ante* regulatory policy.

Another determinant of what might be a weakening of the public administrator's position over time is the fact that incumbents accumulate knowledge capital that can only be acquired by actually operating the facility.

Together with the impossibility of specifying all contingencies in a franchise contract, this implies a potential *ex post* hold-up problem. In other words, private partners maintain the quality of the service, make capacity adjustments or adapt prices only if additional demand by public partners is met. Renegotiation of transport infrastructure investment has been the rule rather than the exception and public concession contract administrators have essentially the same job as regulators.

There is very little disagreement that "competition for the market" does not work as a substitute for real market competition in the provision of transport infrastructure services. Good regulatory institutions are crucial for improving the performance of the natural monopoly segments of the transport sector in general and transport infrastructure in particular. They also form the basis for the introduction of competition into the potentially competitive segments of the transport sector.

2.4. Political economy

A major reason for the disappointment with regulatory policies lies in the fact that regulatory concepts were mostly aimed at increasing efficiency only, while actual policies were fraught with various distributional objectives as well. In the practice of transport policy, distributive objectives sometimes play a greater role than efficiency objectives, even if regulatory policies are not useful instruments for changing the personal or regional distribution of income. The most common redistributive objective of transport policy measures and infrastructure investment has been the broad social aim of giving all of a country's citizens access to transport infrastructure services at "affordable" prices. Other distributive objectives of transport and regulatory policies have been to sustain or increase local employment levels, to achieve a more equal regional income distribution and to favour local equipment manufacturers or construction companies. The latter objectives have at times been argued for by making reference to efficiency objectives, i.e. by arguing the case of realising network economies or increasing returns to scale.

As a consequence, the regulation of natural monopolies has been used as a vehicle to implement a wide range of implicit tax and subsidy programmes. The regulation of transport infrastructure facilities is an effective instrument in achieving distributional objectives for at least two reasons: the monopoly status makes it impossible for the (local) economy to undermine the distributive intentions by behavioural changes or by readjusting business plans; second, the magnitude and nature of the redistribution of income and wealth may be buried in a complex system of tax- and tariff-setting, plus fiscal redistribution which insulates the policy from any meaningful public scrutiny. Such a system of hidden taxes and subsidies is much more difficult to monitor than "on budget" legislative taxes and direct subsidy programmes.

At least to some extent, cyclical changes in the views on regulation have to do with polarised views on the role of government in a market economy. The position favourable to regulation is based on the assumption that it serves a "public interest" by correcting some form of market failure in the particular form of a natural monopoly (Noll, 1989). It corresponds to the metaphor of the political decisionmaker being an omniscient and omnipotent welfare maximiser. It follows that these social welfare maximisers should either directly regulate or manage firms. Such a view emphasizes and emphasized policymaking as a technical problem. It is based on the implicit assumption that once a policy which maximizes or improves social welfare has been found and recommended, it will be implemented as designed and the desired effects will follow. The overwhelming evidence conflicting with this view of planning and policymaking, interpreted as "imperfections" of government decisionmaking, sometimes resulted in the extreme belief that deregulation and the toleration of the associated market imperfections would lead to the best practical outcomes.

Both positions disregard the endogenous process of political decisionmaking; they ignore that a policy proposal is the starting point of a process which is political at every stage -- not only the process of legislation but also the implementation, including the choice or formation of an administrative agency and its subsequent operation. They fall short of incorporating a model of the political process whose essence is that many participants simultaneously try to influence the actions of the immediate policymaker.

Beginning with Stigler (1971) and continuing with Peltzman (1976) and Becker (1983), the public interest view had been challenged by a theory of the regulator's behaviour which implies that compact, well-organised groups (frequently producers) will tend to benefit more from regulation than broadly diffuse groups, like consumers or taxpayers. Regulatory policy will then seek to preserve a politically optimal distribution of rents across the coalition of well-organised groups. As the political benefits (e.g. re-election chances) arise from the redistribution of wealth, the inefficiency created by the regulatory process is limited by the decreased redistribution potential. The discussion at the Round Table, however, showed that a more general recognition of the political process in regulatory policy is needed.

The above-mentioned early literature on "capture" ignores informational asymmetries. In the absence of such asymmetries, firms would be unable to extract rents and therefore have no reason to influence the political process that leads to the regulatory outcomes. Voters and legislators would be able to control their agents (members of committees and agencies) who thus would not get away with policies favouring interest groups over the common good. Only recently have theoretical concepts become available which help to analyse why regulated firms and interest groups have been active in influencing the political process concerning regulation of the provision of transport infrastructure services (Laffont and Tirole, 1993, Part V)[2].

There was strong agreement among the Round Table participants that the political economics of the transport sector explain why deregulation has remained partial in many countries and why the outcomes of deregulation have been unexpected. In contrast to the prescriptions which were developed according to the "public interest" perspective of regulation policy, the segments of the sector which have a potential for competition were not deregulated (long- to medium-distance bus services in some countries, railway transport operations) and, where regulation is needed, often no regulatory agencies were set up. Where regulatory institutions had been set up, regulators almost always lacked independence or were not accountable to the general public.

This leads to the question of whether and to what extent regulators should be given discretionary powers, or their action space be constrained by constitutionally fixed rules. Such rules could, in principle, limit the influence of interest groups, incumbent firms and other well-organised groups and therefore regulatory capture, as is emphasized by the public choice or contractarian literature on regulation (cf. in particular, Buchanan and Tullock, 1962; Buchanan, 1988). If the involvement of interest groups and firms in the political process is aimed at changing the distribution of rents arising from the provision of infrastructure services, this could lead to better regulatory outcomes. However, renegotiation of regulatory arrangements might be required due to the fact that they are incomplete -- that, for example, concession contracts cannot account for all possible contingencies. The restriction on possibilities to renegotiate the terms of regulatory arrangements might well, therefore; increase hold-up problems in an ongoing contractual relationship, i.e. lead to inadequate quality and capacity expansion, due to perceived political risks for the private investors (Hart, 1995). Moreover, it is not certain that the establishment of constitutional rules of regulation does not suffer from a political influence leading to outcomes which differ from the "public interest" prescriptions (Dixit, 1996, Chapter 2).

3. THE EMPIRICAL RECORD OF TRANSPORT SECTOR DECENTRALISATION

Empirical information on the effects of deregulation and privatisation of transport infrastructure provision is scarce and often anecdotal. Reviewing the unsystematic evidence available gives the impression that, overall, deregulation has led to efficiency gains, with substantial differences between the modes of transport. Significant efficiency gains have been achieved where the potential for competition in the market for infrastructure services is relatively high, e.g. for ports or airports.

Systematic international evidence is provided by a survey, "Public Works Financing International" (cf. Estache, 2001 and 2005). It estimated that about 1 137 transport infrastructure projects with private involvement, worth US$684 billion, were planned in 2003. About half of them were toll road projects, a quarter in the railway subsector, and the rest were plans to finance air and sea ports.

The fact that only half of the projects were under construction or in operation by 2003 might indicate that the enthusiasm of the early nineties had suffered in the process of defining the details of public-private projects in practice. The 1997 financial crisis led to a major drop in the number of commitments. At the same time, the dimension of the decrease was greater for poorer countries than for rich countries. The diminished expectations of private investors are reflected in the doubling of the debt-to-equity ratio from the mid-nineties to 2001. The average cost of equity to finance transport infrastructure is, according to World Bank accounts, three percentage points higher than the average of transport infrastructure capital costs.

Another aspect of the diminished expectations is the less-than-expected stability of deregulation and privatisation arrangements: where systematic evidence is available, renegotiation of concession contracts has been the rule rather than the exception. In a sample of infrastructure concession contracts in Latin America, about 30 per cent of all concession contracts which had been set up between 1983 and 2000 were renegotiated. In the transport sector, the share was more than 50 per cent. On average, renegotiations had already started about three years after signing the contracts. Certainly, not all of the renegotiations can be considered to be problematic. On the contrary, the incompleteness of concession contracts makes them inevitable and implies that they may enhance efficiency. The high instability of transport infrastructure concession contracts, however, casts doubts on an optimistic view regarding the optimality of the contracts and the substitutability of competition in the market by competition for the market (Estache, 2005).

The disappointment with the limited extent of political implementation and the achievement of its objectives notwithstanding, efficiency has improved, according to most case studies (Joskow, 1999; Winston, 1993). This was confirmed by the Round Table discussion. Of major importance for the success of deregulation seems to be the time frame of the implementation. Whether privatisation and deregulation should be introduced gradually or by taking one big step is a relevant question without a universal answer. The decisions on the sequencing of reforms have depended on:

- the performance attributes of the existing system and the analysis of where the current performance problems are;
- the complexity of implementing a "big bang" approach, given the imperfections of current regulatory institutions as well as the distortions resulting from the current pricing and taxation regime;
- the capacity of legal and political institutions to support competitive markets for infrastructure services;
- the time required to create market and regulatory institutions; and finally
- the government's capability to commit *ex ante* to a restructuring framework which supports investment and competitive entry.

As has been pointed out above, one of the most important drivers of privatisation and deregulation was the expectation that the provision of transport infrastructure services would be depoliticised and the implementation of hard budget constraints would lead to substantial benefits for the fiscal sector. The evidence which is available so far does not allow a final judgment on whether the fiscal objectives have been achieved: early empirical studies gave a positive picture of the fiscal consequences of privatisation and deregulation. There are, however, indications that the reported effects were largely of a temporary nature. The divestiture of transport infrastructure facilities led to one-off public returns, which led to short-run improvements of the fiscal situation of public authorities. More importantly, the rental payments of private service providers to the public owners have been temporary as well: a recent study looking at debt reductions, sales and rental fee gains has shown that the recurrent expenditures increased significantly. Most of these increases were due to demands or subsidies for passenger transport. The return of operational subsidies might be a signal that the reform of the transport sector has not achieved the objective to depoliticise the sector.

In many cases, the expectation that privatisation and deregulation would do much to increase the level of transport infrastructure investment has not been met either. The share of private capital in financing transport infrastructure has remained relatively low and, with privatisation, the public investment share seems to have dropped by more than what was taken up by private investment. To some extent this has been due to the fact that transport infrastructure investment was a target for realising even more public expenditure cuts than before the reform (Campos *et al.,* 2003b).

Another reason for the limited success of the reforms lies in the fact that in many countries the number of potential bidders for concession or franchise contracts has been very small. Apart from the resulting danger of collusion in the bidding process, this is a potential basis for rent-seeking activities by incumbent private firms.

4. REGULATORY CONCEPTS

To some extent, the disappointment with the regulatory reforms of the transport sector derives from the malfunctioning of the regulatory mechanisms put in place. A last discussion block of the Round Table therefore addressed questions concerning improvement of the regulatory mechanisms.

A regulator's task would be straightforward if he had complete information about the regulated firms' present and future technological and cost opportunities, about the demand of all types of consumers and how they will evolve, the data necessary to pursue distributional objectives and the public's assurance that the regulator can be trusted to pursue his goals efficiently and impartially. The regulator could calculate (second-best) optimal price levels and an optimal tariff structure for every point in time and adjust them as costs, demands and distributional objectives change. The regulator would simply execute a well-defined set of "public interest" objectives efficiently. If this were a useful depiction of reality there would be no reason to separate the regulator from the regulated firm, since the public regulator would have all the information required to produce and price efficiently. This has sometimes been the implicit rather than explicit rationalisation for the public ownership of natural monopoly (sub-) sectors.

The central problem with the design of regulatory mechanisms is the fact that the regulated firm has private information about available technologies and corresponding cost functions, the operating characteristics of its network, the effort it expends to reduce costs, the quality of its services and the responsiveness of its customers to various quality and price signals. Even if this asymmetry of information can be substantially reduced in principle, the potential efficiency gains to be had from such a reduction must be measured against the monitoring and control costs of the regulator.

In fact, standard regulatory concepts did not and do not recognise that cost, demand and quality information is not publicly available and that the monitoring of regulated firms is costly. This holds for "cost plus" contracts where owners or franchisees are allowed to charge prices, or are paid transfers, to cover the full costs plus a "fair" rate of return. It holds for the various pricing rules, the variants of marginal cost pricing and average cost pricing, as well as for rate-of-return regulation. To verify the performance of firms, the regulator has to know not only the firms' actual costs but also their technically possible least costs. If the best practice technology or least cost input-output relationship is unknown to the regulator and/or if the monitoring and verification of the costs and demand information of the regulated firm is expensive, firms do have opportunities to receive information rents. In other words, reported costs might conceal profits such as overstaffing, overinvestment in firms' amenities and a "quiet life", in the sense of a low level of effort to reduce costs or to improve the quality of services provided. In whichever way, the regulator is confronted with the problem of containing informational rents, independently of whether the regulated firm has a public or private owner or is operated by a private concession-taker with public ownership.

The Round Table discussion reflected growing support for the view that private ownership of transport infrastructure facilities poses enormous problems to achieve regulatory objectives. At the same time, there was broad agreement that public ownership does not obviate or reduce the need for regulation. The rent-seeking opportunities of public firms and/or private franchisees are not fundamentally different from those of regulated privately owned and operated firms. Two main arguments seem to favour public ownership of transport infrastructure facilities:

- It is often easier for private firms to hide information from the regulator. The disclosure of certain relevant business information might conflict with the interests of private shareholders;

- The high degree of indivisibility of transport infrastructure investment and the fact that markets for infrastructure facilities do not exist imply that there is a hold-up problem: private investors recognise the risk that, after the investment has been realised and expenditures are sunk, governments might revise initial commitments to the investors' disadvantage. Due to these circumstances, the chances of combining private ownership of transport infrastructure facilities with a strong regulatory regime seem to be limited.

Against this backdrop, the regulatory concepts have to be evaluated according to their effectiveness in reducing the informational asymmetries between the regulator and the monopolistic firm, and the associated potential for earning informational rents for the regulated firm.

Yardstick competition is perhaps the most important instrument allowing the regulator to induce the regulated firms to reveal truthful cost information. The basic concept foresees that the regulated firm sets its price equal to the average of the marginal costs of other similar firms producing the same good or service. Additionally, the firm receives a lump-sum transfer, equal to the average investment of other firms to reduce their costs. Thus, for each firm a "yardstick" is defined by the performance of other firms. Shleifer (1985) has shown that, in a non-cooperative equilibrium, each firm has an interest in revealing its true costs and investment to reduce its costs.

Although yardstick competition has been applied in the Japanese railway sector (Okabe, 2004) and in the Norwegian bus industry (Dalen and Gomez-Lobo, 2003c), it has not been introduced in the transport infrastructure sector beyond the proposal of extending benchmarking efforts to yardstick competition (Estache *et al.,* 2002, on Mexican port liberalization).

There are four major reasons for the slow progress in introducing yardstick competition in the provision of transport infrastructure services:

- A first reason lies in the fact that many infrastructure facilities are interdependent or form a network. If there are unbounded network economies, the introduction of yardstick competition implies that the country or region in question has to forego network economies to increase the (surrogate) competitive pressure. For railways, for example, this seems to be a difficult decision to take, as the British example shows.

- The benefits of yardstick competition and performance evaluation can only be reaped if the agents act non-cooperatively. As Shleifer (1985) had already remarked, "an important limitation of yardstick competition is its susceptibility to collusive manipulation (p. 327)." The stronger the network economies, the smaller will be the number of firms subject to yardstick competition. Given a small number of firms and repeated interaction between the regulated firms and the regulator, collusion between the regulated transport infrastructure service providers is a significant possibility. The danger of collusion can be reduced by changing the yardstick scheme but at the cost of achieving a social optimum. This has recently been confirmed by experimental research (Potters *et al.*, 2003). What is more, collusion-proof yardstick mechanisms appear to be hard to define in general (Laffont and Martimort, 2000).

- Even if competing transport infrastructure entities of the same mode can be defined, the question arises whether they could and would have identical cost functions in a state of technical efficiency. Transport infrastructure facilities, as local monopolies, are strongly influenced by exogenous factors such as geography, as well as population density and its distribution. Therefore, firms subject to yardstick competition will be heterogeneous and it may be difficult to account for the heterogeneity in defining the correct, firm-specific yardstick (Bouf and Leveque, 2005).

- The heterogeneity problem is made more acute if yardstick competition is implemented in a system of decentralised regulatory powers, entailing a danger of the emergence of differing local yardsticks, leading to discrimination between firms (Bivand and Szymanski, 1997).

For all these reasons, the chances for implementing yardstick competition for maintenance may be greater than for construction and capacity adjustment of transport infrastructure. A first important step towards the implementation of yardstick competition is the adoption of accounting conventions which allow benchmarking between firms using comparable data at a fairly aggregate level. Price cap regulation, together with the assignment of the burden of proof of cost conditions and investment levels to infrastructure service providers, were seen as an important step toward improving existing regulatory regimes.

5. SUMMARY OF MAIN POLICY CONCLUSIONS

While deregulation and privatisation in the transport sector has led in general to increases in productivity, not all the desired effects of the reforms have materialised. This holds in particular for transport infrastructure investment, where privatisation and deregulation have not caused the expected mobilisation of private resources and where franchise relations have not been as stable as expected. Based on the current conceptual discussions and a review of the reform results, the Round Table drew the following conclusions:

- The internal heterogeneity of the transport sector does not allow for polar policy prescriptions such as "deregulate and privatise" or "tax finance and produce publicly". Whenever competition in the market is possible, and entry and exit costs admit potential competition, market outcomes will be superior to detailed regulatory regimes and public production.

- There are, however, major parts of the industry where indivisibilities, network economies and the absence or malfunctioning of resale markets for investment goods lead to inefficient market outcomes. Strict criteria should be applied to identify those sections which require any regulation at all.

- In these sections the opportunities for private ownership, in particular for transport infrastructure, have sometimes been overrated. The regulatory tasks for the monopolistic parts of the sector do not differ fundamentally between different ownership regimes.

- The transaction costs induced by regulatory regimes deserve greater attention than in the past. A major area of debate in this respect concerns the separation of ownership of infrastructure from transport operations. The Round Table discussion reflected a growing concern that the neglect of transaction costs has led to the problematic consequence of vertical disintegration.

- Any regulatory policy has to acknowledge informational asymmetries between the actors involved. Some of the traditional regulatory concepts have ignored the enormous monitoring and control costs incurred by regulators.

Transport sector reforms did not have much success either in depoliticising the provision of transport infrastructure services, or transport services in general. In most cases, regulators do not enjoy the independence envisaged at the beginning of the reform process. The incompleteness of concession contracts and contract partners' mutual commitment problems have led to frequent renegotiations with a political character.

Both these characteristics, the information problem and the lack of depoliticising regulatory policies, suggest that rule-bound, performance-based mechanisms such as yardstick competition should play a stronger role for the transport sector. The implementation of such mechanisms would reduce the problems of information for regulators and the chances for discretionary, opportunistic behaviour by regulatory agencies.

NOTES

1. This problem has attracted much attention in location theory and more recently has been studied for the competition between airlines (Button, 1999, 2003).

2. A more general review of the analytics of the political decisionmaking process can be found in Tirole (1994) and Dixit (1996). Grossman and Helpman (2001, 2002) have extended this approach by developing an analytical framework of interest group behaviour with special reference to trade policy. Wilson (1980, 1989) studies the dynamics of newly created agencies.

BIBLIOGRAPHY

Baumol, W.J. (1982), Contestable markets: An uprising in the theory of industrial structure, *American Economic Review,* 72: 1-15.

Baumol, W.J., J.C. Panzar and R.D. Willig (1982), *Contestable Markets and the Theory of Industry Structure,* San Diego.

Becker, G. (1983), A theory of competition among pressure groups for political influence, *Quarterly Journal of Economics,* 98: 371-400.

Bivand, R. and S. Szymanski (1997), Spatial dependence through local yardstick competition: Theory and testing, *Economics Letters,* 55: 257-265.

Bouf, D. and J. Leveque (2005), Yardstick competition for transport infrastructure services, in: *Round Table 129*, *The Limits of (De-) Regulation of Transport Infrastructure Services,* OECD/ECMT, Paris.

Buchanan, J.M. (1988), Contractarian political economy and constitutional interpretation, *American Economic Review (Papers and Proceedings),* 78: 135-139.

Buchanan, J.M. and G. Tullock (1962), *The Calculus of Consent,* Ann Arbor.

Button, K.J. (1999), Shipping alliances: Are they at the "core" of solving instability problems in shipping?, in: *Liner shipping: what's next?,* Proceedings of the 1999 Meeting of the International Association of Maritime Economists, Halifax.

Button, K.J. (2003), Does the theory of the "core" explain why airlines fail to cover their long-run costs of capital?, *Journal of Air Transport Management,* 9: 5-14.

Campos, J., A. Estache, N. Martin and L. Trujillo (2003), Macroeconomic effects of private sector participation in infrastructure, in: W. Easterly and L. Serven (eds.), *The Limits to Stabilization -- Infrastructure, Public Debts and Growth in Latin America,* Palo Alto.

Dalen, D.M. and A. Gomez-Lobo (2003), Yardstick competition on the road: Regulatory contracts and cost efficiency in the Norwegian bus industry, *Transportation,* 30: 371-386.

Davidson, C. and R. Deneckere (1986), Long-run competition in capacity, short-run competition in price and the Cournot model, *Rand Journal of Economics,* 17: 404-415.

Demsetz, H. (1968), Why regulate utilities?, *Journal of Law and Economics,* 11: 55-66.

Dixit, A.K. (1996), *The Making of Economic Policy,* Cambridge, Mass.

Estache, A. (2001), Privatization and regulation of transport infrastructure in the 1990s, *World Bank Research Observer,* 16: 85-107.

Estache, A., M. Gonzales and L. Trujillo (2002), Efficiency gains from port reform and the potential for yardstick competition: Lessons from Mexico, *World Development,* 30: 545-560.

Estache, A. and T. Serebrisky (2005), Where do we stand on transport infrastructure deregulation and public-private partnership?, in: *Round Table 129*, *The Limits of (De-) Regulation of Transport Infrastructure Services,* OECD/ECMT, Paris.

Grossman, G.M. and E. Helpman (2001), *Special Interest Politics,* Cambridge, Mass.

Grossman, G.M. and E. Helpman (2002), *Interest Groups and Trade Policy,* Princeton.

Hart, O. (1995), *Firms, Contracts and Financial Structure,* Oxford.

Joskow, P.L. (1999), Regulatory priorities for reforming infrastructure sectors in developing countries, in: B. Pleskovic (ed.), *Annual World Bank Conference in Developing Economics 1998,* Washington, DC.

Knieps, G. (2005), The limits of competition in transport markets, in: *Round Table 129*, *The Limits of (De-) Regulation of Transport Infrastructure Services,* OECD/ECMT, Paris.

Kreps, D. and J. Scheinkman (1983), Quantity precommitment and Bertrand competition yield Cournot outcomes, *Bell Journal of Economics,* 14: 326-337.

Laffont, J.-J. and D. Martimort (2000), Mechanism design with collusion and correlation, *Econometrica,* 68: 309-342.

Laffont, J.-J. and J. Tirole (1993), *A Theory of Incentives for Procurement and Regulation,* Cambridge, Mass.

Nash, C.A. and J.P. Toner (1999), Competition in the railway industry, *Journal of Competition Law and Policy,* 1: 197-222.

Noll, R.G. (1989), Economic perspectives on the politics of regulation, in: R. Schmalensee and R. Willig (eds.), *Handbook of Industrial Organization*, Vol. II, Amsterdam.

Okabe, M. (2004), New passenger railway fares, *Japan Railway & Transport Review,* 37: 4-15.

Peltzman, S. (1976), Towards a more general theory of regulation, *Journal of Law and Economics,* 19: 211-40.

Ponti, M. (2005), Theoretical and policy issues, in: *Round Table 129, The Limits of (De-) Regulation of Transport Infrastructure Services,* OECD/ECMT, Paris.

Potters, J., B. Rockenbach, A. Sadrieh and E.E.C. van Damme (2003), Collusion under yardstick competition. An experimental study, CentER Discussion Paper 2003-97, Tilburg University, Tilburg.

Shleifer, A. (1985), A theory of yardstick competition, *Rand Journal of Economics,* 20: 41-58.

Stigler, G.J. (1971), The theory of economic regulation, *Bell Journal of Economics and Management Science,*2: 3-21.

Stiglitz, J.E. (1987), Technological change, sunk costs, and competition, in: J.E. Stiglitz, D. McFadden, and S. Peltzman (eds.), *Brookings Papers on Economic Activity*, Vol. 3, Special Issue on Microeconomics.

Tirole, J. (1994), The internal organisation of government, *Oxford Economic Papers,* 46: 1-29.

Wilson, J.Q. (1980), *The politics of regulation,* New York.

Wilson, J.Q. (1989), *Bureaucracy: What Government Agencies Do and Why They Do It,* New York.

Winston, C. (1993), Economic deregulation: Days of reckoning for microeconomists, *Journal of Economic Literature,* 31: 1263-1289.

LIST OF PARTICIPANTS

Prof. Emile QUINET **Chair**
Head of Department
Ecole Nationale des Ponts et Chaussées
Département d'Economie et des Sciences
28 rue des Saints-Pères
F-75007 PARIS
France

Mr. Dominique BOUF **Co-rapporteur**
Laboratoire d'Economie des Transports (LET)
MRASH
14 avenue Berthelot
F-69363 LYONS Cedex 07
France

Mr. Julien LEVEQUE **Co-rapporteur**
Laboratoire d'Economie des Transports (LET)
MRASH
14 avenue Berthelot
F-69363 LYONS Cedex 07
France

Mr. Antonio ESTACHE **Rapporteur**
World Bank
1818 H Street, NW
WASHINGTON DC 20433
USA

Prof. Dr. Günter KNIEPS **Rapporteur**
Director
Albert-Ludwigs-Universität Freiburg
Institut für Verkehrswissenschaft und Regionalpolitik
Kollegiengebäude II
Platz der Alten Synagoge
D-79085 FREIBURG
Germany

Prof. Marco PONTI **Rapporteur**
President
TRT Trasporti e Territorio SRL
Via Rutila, 10/8
I-20146 MILANO
Italy

Prof. Dr. Ulrich BLUM
Technische Universität Dresden
Fakultät Wirtschaftswissenschaften
Schumannbau C264
Münchner Platz
D-01069 DRESDEN
Germany

Prof. Sergio BOLOGNA
Consultant
Associazione Italiana di Logistica et di Supply Chain Management
Via Casale 7
I-20144 MILAN
Italy

Mr. Alim DEMCHUK
Head of Economic Analysis and Strategic Planning Department
Ministry of Transport of Ukraine
14 av. Peremogiy
UKR-01135 KIEV
Ukraine

Mr. Oleksandr NASTYCH
Chief Expert, Foreign and Economic Relations Department
Ministry of Transport of Ukraine
14 av. Peremogiy
UKR-01135 KIEV
Ukraine

Mr. John DODGSON
Director and Head of European Transport Practice
National Economic Research Associates - NERA
15 Stratford Place
GB-LONDON, W1C 1BE
United Kingdom

Mr. Jeremy DREW
Drew Management Consultants
63 Aberdeen Road
GB-LONDON N5 2XB
United Kingdom

Mrs. Olga KRISTOFIKOVA
Counsellor
Ministry of Transport
Transport Policy Department
P.O. Box 9
Nabrezi Ludvika Svobody 12
CZ-110 15 PRAGUE 1
Czech Republic

Mrs. Celina LUIS
Director
Institut National de Transport Ferroviaire (INTF)
rue Padre Luis Aparicio, n° 7
P-1150 248 LISBON
Portugal

Prof. Rico MAGGI
Università della Svizzera Italiana
Istituto di ricerche economiche (IRE)
Via Maderno 24
C.P. 4361
CH-6904 LUGANO
Switzerland

Prof. Andrés MONZON
Escuela Tecnica Superior de Ingenieros de Caminos
Transport Department - UPM
Ciudad Universitaria, s/n
E-28040 MADRID
Spain

Prof. Christopher NASH
University of Leeds
Institute for Transport Studies
36 University Road
GB-LEEDS, LS2 9JT
United Kingdom

Dr. John PRESTON
Transport Studies Unit
University of Oxford
11 Bevington Road
OXFORD OX2 6NB
United Kingdom

Ing. Andrea RICCI
Istituto di Studi per l'integrazione dei sistemi
Via Flaminia 21
I-00196 ROME
Italy

Prof. Dr. Wlodzimierz RYDZKOWSKI
Chairman of Department
University of Gdansk
Department of Transportation Policy
Armii Krajowej 119/121
PL-81-824 SOPOT
Poland

Mr. Toshiake SAKATSUME
EBRD
Office of the Chief Economist
1 Exchange Square
GB-LONDON EC2A 2JN
United Kingdom

Mr. B.E. SPIERING
Ministry of Transport, Public Works and Water Management
Rijkswaterstaat
P.O. Box 20906
NL-2500 EX THE HAGUE
Netherlands

Prof. Lourdes TRUJILLO
Universidad de Las Palmas de Gran Canaria - General
Campus de Tafira
Departamento de Analisis Economico Aplicado
E-35017 LAS PALMAS
Spain

JOINT OECD/ECMT TRANSPORT RESEARCH CENTRE

SECRETARIAT:

Dr. Andreas KOPP
Chief Economist

Dr. Michel VIOLLAND
Administrator

Mrs. Julie PAILLIEZ
Assistant

Ms. Françoise ROULLET
Assistant

ALSO AVAILABLE

What Role for the Railways in Eastern Europe? Series ECMT – Round Table 120 (2002)
(75 2002 04 1 P) ISBN 92-821-1371-X

Managing Commuters' Behaviour - A New Role for Companies. Series ECMT – Round Table 121 (2002)
(75 2002 11 1 P) ISBN 92-821-1299-3

Transport and Exceptional Public Events. Series ECMT – Round Table 122 (2003)
(75 2003 04 1 P) ISBN 92-821-1305-1

Vandalism, Terrorism and Security in Urban Passenger Transport. Series ECMT – Round Table 123 (2003)
(75 2003 07 1 P) ISBN 92-821-0301-3

Transport and Spatial Policies: The Role of Regulatory and Fiscal Incentives. Series ECMT – Round Table 124 (2004)
(75 2004 09 1 P) ISBN 92-821-2321-9

European Integration of Rail Freight Transport. Series ECMT – Round Table 125 (2004)
(75 2004 06 1 P) ISBN 92-821-1319-1

Airports as Multimodal Interchange Nodes. Series ECMT – Round Table 126 (2005)
(75 2005 03 1 P) ISBN 92-821-0339-0

16th International Symposium on Theory and Practice in Transport Economics. 50 Years of Transport Research (2005)
(75 2005 05 1 P) ISBN 92-821-2333-2

Time and Transport. Series ECMT – Round Table 127 (2005)
(75 2006 01 1 P) ISBN 92-821-2330-8

National Systems of Transport Infrastructure Planning. Series ECMT – Round Table 128 (2005)
(75 2005 10 1 P) ISBN 92-821-2341-3

OECD PUBLICATIONS, 2, rue André-Pascal, 75775 PARIS CEDEX 16
PRINTED IN FRANCE
(75 2006 02 1 P) ISBN 92-821-2345-6 – No. 54927 2006

129

TRANSPORT SERVICES: THE LIMITS OF (DE)REGULATION

While deregulation and privatisation in the transport sector have led to increases in productivity in general, not all reform hopes have materialised. In particular, the reform of the provision of infrastructure services has not caused the expected mobilisation of private resources, and concession relations have been less stable and less efficiency-enhancing than expected. In view of current discussions of reform results, the Round Table focused on the following issues:

Where are the limits for deregulation?
The discussion identified the conditions under which competition and potential competition can be expected to work. More care has to be applied to single out the transport sub-sectors where these conditions hold.

Which are the crucial factors that necessitate regulation?
Many parts of the transport sector are fraught with indivisibilities, network economies, sector-specific assets or lack of resale markets for investment goods. Where these factors play an important role, regulation might improve the efficiency of the transport system.

What is the role of the transaction costs of regulation?
The neglect of (surrogate) market transaction costs, in particular in the case of vertical disintegration, has led to lower than expected benefits from the reforms.

What is the cost of regulation?
Regulatory policies have to take account of the information asymmetries between the actors involved. Monitoring and control costs have often prohibited the depoliticising of regulatory processes. The Round Table discussed to what extent a rule-bound, performance-based regulation could contain the friction resulting from discretionary regulatory powers.

(75 2006 02 1 P1) ISBN 92-821-2345-6