WORLD BANK INSTITUTE

Promoting knowledge and learning for a better world

20742

June 2000

Other WBI Development Studies
(In order of publication)

Principles of Health Economics for Developing Countries
William Jack

Chile: Recent Policy Lessons and Emerging Challenges
Edited by Guillermo Perry and Danny M. Leipziger

Curbing Corruption: Toward a Model for Building National Integrity
Edited by Rick Stapenhurst and Sahr J. Kpundeh

Resetting Price Controls for Privatized Utilities: A Manual for Regulators
Richard Green and Martin Rodriguez Pardina

Preventing Bank Crises: Lessons from Recent Global Bank Failures
Edited by Gerard Caprio, Jr., William C. Hunter, George G. Kaufman, and Danny M. Leipziger

Enterprise Restructuring and Unemployment in Models of Transition
Edited by Simon Commander

Poverty in Russia: Public Policy and Private Responses
Edited by Jeni Klugman

Enterprise Restructuring and Economic Policy in Russia
Edited by Simon Commander, Qimiao Fan, and Mark E. Schaffer

Infrastructure Delivery: Private Initiative and the Public Good
Edited by Ashoka Mody

Trade, Technology, and International Competitiveness
Irfan ul Haque

Corporate Governance in Transitional Economies: Insider Control and the Role of Banks
Edited by Masahiko Aoki and Hyung-Ki Kim

Continued on inside back cover

WBI Development Studies

Privatization and Regulation of Transport Infrastructure

Guidelines for Policymakers and Regulators

Edited by
Antonio Estache
Ginés de Rus

The World Bank
Washington, D.C.

First printing June 2000

The World Bank Institute (formerly the Economic Development Institute) was established by the World Bank in 1955 to train officials concerned with development planning, policymaking, investment analysis, and project implementation in member developing countries. At present the substance of WBI's work emphasizes macroeconomic and sectoral policy analysis. Through a variety of courses, seminars, and workshops, most of which are given overseas in cooperation with local institutions, WBI seeks to sharpen analytical skills used in policy analysis and to broaden understanding of the experience of individual countries with economic development. Although WBI's publications are designed to support its training activities, many are of interest to a much broader audience.

This report has been prepared by the staff of the World Bank. The judgments expressed do not necessarily reflect the views of the Board of Executive Directors or of the governments they represent.

The backlist of publications by the World Bank is shown in the annual Index of Publications, *which is available from the Office of the Publisher.*

Antonio Estache is principal economist in the World Bank Institute's Governance, Regulation, and Finance Division.

Ginés de Rus is a professor at the Universidad de las Palmas de Gran Canaria in Spain.

For information on the Regulatory Reform and Competition program at the World Bank Institute, visit: http://www.worldbank.org/wbi/regulation/books.htm.

Library of Congress Cataloging-in-Publication Data

Privatization and regulation of transport infrastructure: guidelines for policymakers and regulators/edited by Antonio Estache, Ginés de Rus.
p. cm. — (WBI development studies)
Includes bibliographical references.
ISBN 0-8213-4721-7
1. Transportation--Planning. 2. Transportation and state. 3. Transportation--Deregulation. 4. Privatization. 5. Competition. I. Estache, Antonio. II. Rus, Ginés de. III. Series.

HE193.P73 2000
388--dc21 00-036493

Contents

Boxes

Figures

Tables

Foreword

The provision of transport services has changed dramatically during the last two decades. At the end of the 1970s, most countries relied on the public sector both to produce transport services and to build their basic infrastructure, namely, airports, roads, railways, and ports. The role of private firms in transport was secondary, and governments tackled the main tasks.

This arrangement has been turned upside down. After almost two decades of privatization, the private sector has now become the main actor providing many aspects of transport infrastructure and services. Publicly owned companies have been sold, and from Asia to Latin America, many transport services have been concessioned to private operators. Remarkably, the private sector is also starting to build and finance the development of basic infrastructure, although public sector financing will continue to be important, especially in the road sector and in activities that carry strong social implications.

Getting the private sector involved may be the easy part of transforming the sector. Having governments effectively take on their new role as regulators may be the toughest challenge. For instance, in some cases a dominant firm may use its powers not only to raise prices, but to deter entry, and government policies must be designed so as to restrict predatory actions and restraints to trade. The objective of this book is to help governments learn about this new role.

The issues discussed are timely. In many countries private operators have now become a critical mass, and regulators need to take effective action if private participation is to yield sustainable, efficient, and fair outcomes in the transport sector. Transport sector regulators must learn how to promote competition to obtain low fares and efficient services and how to safeguard users' interests when competition is weak or nonexistent. A major challenge in this connection is to strike a balance between retaining public ownership of the infrastructure assets and promoting efficient construction and operation of such assets by private firms.

Joseph E. Stiglitz
Professor of Economics, Stanford University

Acknowledgments

This work has benefited from comments by and the experiences of many participants at a course on transport regulation organized several times by the World Bank Institute. Our thanks to all of them. We hope that this book will help them and other transport regulators in their daily tasks. Please note that we are most interested in receiving feedback from readers and learning more about experiences with transport regulation around the world that could be included in a future edition.

We would also like to express our gratitude to several colleagues who read earlier versions of the manuscript and provided us with insightful comments, especially Ian Alexander, Nils Bruzelius, José Carbajo, Darius Gaskins, José A. Gómez-Ibáñez, Andrés Gómez-Lobo, José L. Guasch, Jan Hoffman, José L. Irigoyen, Marc Juhel, Alfred H. Nickesen, Emile Quinet, and Louis S. Thompson.

The editorial skills of John P. Didier, International Communications, Inc., and Erin Farnand helped tremendously in polishing the chapters and giving the book its present shape. Any remaining errors or omissions are, of course, our exclusive responsibility.

Antonio Estache
World Bank Institute
Aestache@worldbank.org
Tel: +1 (202) 458 14 42
Fax: +1 (202) 334 83 50

Ginés de Rus
Universidad de Las Palmas de Gran Canaria
gderus@empresariales.ulpgc.es
Tel: + 34 928 45 18 08
Fax: + 34 928 45 81 83

1

Introduction

The 1990s saw a dramatic increase in the liberalization of transport policies and a strengthening of the role of private operators and investors in transport infrastructure worldwide. This increased private sector participation has often reflected changing ideologies about the role of the state and dissatisfaction with publicly provided services. The main driving force behind it, however, generally has been the pressure to look for private financing imposed on governments by lasting fiscal crises. This change in the financing of the sector is also providing an opportunity to restructure it in an attempt to improve its efficiency and sustain these improvements.

In the minds of many policymakers, this search for sustained improvement in efficiency is probably secondary to the need to find additional financing, but it is at the core of the new role of the government in transport. Indeed, in addition to the important responsibilities of defining policies and strategies, monitoring safety, and financing some of the less attractive segments of the sectors, for example, rural and secondary roads, governments must also now be ready to become fair economic regulators of many of the privately operated transport services and infrastructures. The restructuring process often creates new monopolies or oligopolies in which the price, investments, and service quality commitments of operators must be supervised to protect transport users. Moreover, fair regulation is also needed because these operators have rights and the government must be held accountable to the commitments it makes to them as part of the restructuring process.

To implement this economic regulation, most reforming countries are creating new regulatory agencies or units. These regulatory bodies must be effective to ensure that prices are neither excessive nor inadequate, that services meet the desired standards, and that governments and investors comply with the commitments they make. To be as effective as possible, these concerns must be addressed as part of the overall organization of sector reform.

First, the regulatory concerns must be addressed during the privatization stage.[1] Ideally, the future regulators should be involved in the preparation of the sector reform and ensure that their regulatory needs are built in as part of the obligations imposed on the operators during the transfer of responsibilities. Regulators must at least understand the consequences of the restructuring and contract design choices made by the privatization teams. These choices define the constraints and limitations as well as the opportunities that the regulators have to interact with the regulated companies of the sector.

Second, to ensure fairness in regulatory decisions, regulatory bodies that are independent from political interference, but that are also held accountable for their decisions, should take on regulatory concerns. Just as important, regulatory staff members with sufficient skills to be fair while making the most of their autonomy should address regulatory concerns. This is why they must understand the regulatory options available to regulate prices and quality. They must also understand the minimum set of indicators that are used not only to monitor the performance of the regulated companies, but also to increase the transparency of the monitoring process, and hence the accountability of the regulators.

Unfortunately, the government's transition into this new role generally is proving to be more challenging than anticipated, requiring significant adjustments to ensure that efficiency and financial gains are achieved. When it comes to regulation, public sector governance often is weak. In some countries, "regulatory capture," a process in which the regulatory body ends up identifying mostly with the concerns of the industry, is rampant.[2] In other cases, excessive government interference with the regulatory process has amounted to what some would argue is partial expropriation. This poor regulatory governance is becoming a concern for all governments, for the following reasons. First, the international financial crises of 1998–99 have made potential investors much more aware of the risks involved in investing in

---

1. Throughout this book, the concept of privatization is used in a wide sense and includes many types of public-private collaborations that do not require any change in ownership of assets. Most so-called privatization transactions that have taken place in the transport sector are concession contracts, as explained in chapter 2.

2. Although consumers or labor can also "capture" agencies.

sectors that are sensitive to the overall macroeconomic economic situation, and the transport sector is clearly one in which a good share of demand is derived from overall economic activity. More than ever, privatization teams and regulators must ensure that investors receive fair treatment and returns on their investment. If investors lack that confidence, they are very likely to demand higher returns to hedge against regulatory risk or further increase the risk that they will simply not choose to invest enough.

Second, governments often are worried about the emotional load surrounding many privatization processes and are aware of the need for protection from excessive or abusive prices or declining quality of service. This is why it is just as important to ensure that the regulatory regime and process will allow users to enjoy visible gains from sector reforms. Moreover, the regulatory process must be designed to allow these users to voice their concerns through formal channels rather than leaving them to informal channels that are more susceptible to manipulation by political or business interest groups.

Increasingly, governments are also recognizing that one of the main reasons for this poor regulatory governance is that the civil servants recruited to staff these agencies do not have the necessary technical skills to transform them into effective *economic* regulators. Demand for training and training materials to develop regulatory skills in the transport sector is unmet. The main purpose of this book is to contribute to the development of these regulatory skills. To that effect, the book takes stock of what practitioners and academics know about the major challenges that governments are likely to face in taking on their new role in transport.

This book has two parts. The first consists of chapter 2, which provides an overview of what economic theory has to say about why economic regulation is important. Its objective is to introduce potential regulators to some of the key underlying concepts. It may seem too basic for most regulators with recent graduate training in economics, but it may be a useful overview for many of the other professionals who can be recruited to serve as the staff of regulatory bodies. It provides theoretical support to the sector-specific chapters that constitute the second part of the book.

The second part covers four subsectors: airports, ports, railways, and roads. Each chapter can stand apart from the rest of the book and be read on its own, but to facilitate comparisons across subsectors, they all follow exactly the same structure. The first section provides a snapshot of the key economic characteristics of the sector and discusses their relevance from the viewpoint of a regulator. The second section summarizes the main privatization and regulation trends that have been observed in the sector.

The idea is to give an overview of the main options offered by international experience and to cover a few case studies that illustrate these options. The third section covers price regulation and highlights the price-related issues that characterize the sector. The fourth section does the same for quality regulation. The fifth section discusses the main performance indicators that the sector's regulators should be able to rely on to be effective in their jobs.

Because space is limited, we have not been able to address every issue. One key omission is the recognition that many viewpoints are represented in the privatization process and that the outcomes often reflect the biases of the privatization team heads. In an attempt to address this omission, and despite the many common elements, we wrote each chapter with a somewhat different emphasis. In its discussion of airports, chapter 3 emphasizes the restructuring options and their consequences for the various regulatory issues. With ports, chapter 4 emphasizes contract design and its importance for a regulator. Chapter 5 analyzes the rail industry and highlights many of the issues relating to longer-run competition and the strategic needs of the sector, such as intermodal competition and access pricing. Chapter 6 explores roads from a clear project finance viewpoint and tries to show why and how regulators can enhance project finance contracts to ensure that the longer-run regulatory needs are covered.

2

The Regulation of Transport Infrastructure and Services: A Conceptual Overview

Antonio Estache and Ginés de Rus

Transport services have traditionally been subject to tight economic regulation with respect to entering and exiting the market, as well as with respect to the quality of prices and services. In many countries, the public sector itself has traditionally designed, built, and operated road and rail networks and airport and port systems, which is the ultimate form of regulation. This type of government intervention has generally resulted in excessive costs that are not usually matched by prices or quality, creating an outcome that reflects the interests of the sector's civil servants of contractors, unions, and other interest groups more than the preferences of the users and taxpayers. In addition, political interference in pricing and employment decisions has also often resulted in public deficits in the sector. Because these deficits are increasingly difficult to finance through government resources, service rationing has become common, despite growing demand and willingness to pay for more and better services. These are some of the problems that lead to the changes in the sector. These changes, however, do not imply that the government is no longer needed. The government does have to take on new responsibilities. This chapter discusses the theoretical justification for a government role in transport and explains when and how the new responsibilities in economic regulation are justified.

Why Economic Regulation of Transport?

Most countries continue to justify strong public intervention in transportation for specific local, political, and strategic reasons despite frustrations with the quality of service offered by public operators. Generally, governments make this justification out of the need to guarantee access to markets for both goods and populations. This is quite legitimate, because it is the main way that lagging regions can catch up with the faster-growing regions of a country. Governments are also concerned about national security, which is historically why airports and ports have often been under tight military control in many countries.

While these justifications for a strong government presence have a value on their own, many governments also argue the demand-side and supply-side economic foundations of their involvement. On the demand side, many governments have long recognized that transport users are exposed to serious risks because transport of goods and people has few substitutes, even imperfect ones. In technical jargon, the demand for transport service is captive to the operators. The fear that these captive users could be exposed to abuses by uncontrolled, monopolistic service providers has been the most common justification for government intervention. While this risk clearly justifies some type of government action, it does not necessarily imply that the government has to take over the provision of the service.

The most highly demanded transport activities also are politically sensitive for governments, but no reason exists in principle for the private sector not providing them. Their characteristics tend to make them what economists call "private goods." For a transport service or infrastructure to be a private good, it must fulfill two conditions. First, the excludability condition: the operators should be able to exclude from the service or the use of the infrastructure potential users who are not willing to pay for the service or access to the infrastructure. Second, the rivalry condition: the operators should not leave the other users indifferent because it may result in a deterioration of service or may exclude potential users. Bus services, congested toll roads, ports, and airport infrastructures all meet these two aspects of this definition because they are in high demand. A rural road does not meet these criteria because it lacks demand. In general, setting up a tamperproof system to recover tolls on these roads would be costly, although with technological improvements, it should be a realistic option in the foreseeable future. Until then, due to the lack of enforceability and rivalry, they will remain a prime candidate for public sector responsibility—although not necessarily a provision—as will many of the activities for which exclusion and/or rivalry are difficult to obtain or for which demand is financially unjustified for private investors.[1]

If these distinctions have been well understood in the academic world for some time, their implications for the specific role of the government in the transport sector are only recently coming to light. Until not too long ago, policymakers did not appreciate that under some circumstances, the private sector can be the main provider—and financier—of a transport service or infrastructure, and that the participation of the private sector can be structured in many different ways (see box 2.1 for a brief summary of the various organizational forms of transport services). To a large extent, the recent recognition of the potential new role of the private sector in transport stems from technological changes that have altered the nature of the supply side of the transport market, although not evenly across modes.

Transport activities generally present characteristics that influence the structure of a specific transport market. One of the characteristics most commonly used to justify a strong government role or public monopoly are the economies of scale generally assumed to prevail in the sector; that is, the average costs are "always" decreasing as the volume of traffic operated by a firm increases, implying that it will make sense to have a single operator running all of the traffic, and that provision by a single firm will always be cheaper than by more firms. For many transport activities, this makes more sense at first sight than after more scrutiny. Much more room is available for multiple players than common wisdom would suggest among practitioners. Airports, for instance, have scale economies in landing operations and scale dis-economies in the handling of passengers at the terminal, which can be addressed through more competition. The larger the airport, the longer the passengers have to wait to get service. Similarly, in ports the average cost per dock declines with the volume of traffic. This suggests economies of scale, but when considering waiting times there is a clear role for multiple docks.

Recognizing that the possibility of unbundling transport services provides more scope for competition in the sector is critical. It implies that the initial rationale for a single provider progressively disappears. This

1. Conceptually, many of these goods are not strictly private goods, but what economists call club goods. The service provider can exclude users from joining the club (in other words, entering the bus, taking the plane, or driving on the road), but the optimal size of the club is, in general, relatively large to make the business worthwhile. Few people have their own road and this is why the government will continue to be in the road business.

Box 2.1. *Organizational Forms for the Delivery of Transport Activities*

A variety of options are available in between the strictly public or strictly private operations of the sector resulting from a divestiture of assets or from a build-operate-transfer or similar contract for greenfield projects. They are differentiated by the distribution of responsibility for the various aspects of the business (management/operations and investment) and for the commercial risks associated with this business between the public sector and the private sector.

- *Programs/performance contracts.* These contracts are an agreement between an autonomous public enterprise and the ministry or agency with which it is affiliated. The managers of the public enterprise commit to specific objectives, generally output targets, productivity indicators, or costs cuts, within a specific period of time. These contracts tend to be quite short—two to five years—and renewable. Payments to the public enterprise are generally through subsidies to finance investment needs, seldom to operations. They generally fail to reach their goals in the medium and long run. Experience suggests that this stems from the political temptation to interfere with the management of public enterprises in sensitive sectors. This is why their use is declining in developing countries.
- *Management contracts.* The assets of the transport company continue to be public, but operational management becomes private. The private operator is paid through a fee (generally a fixed component plus a success fee depending on the revenue from the business) and is not responsible for either investment or commercial risk. This has the advantage of bringing in private management skills, and any associated innovations, for a period of two to five years. This should also be seen as a transitional solution because, from a fiscal point of view, it is not attractive as the government continues to take on all risk and finances all investments.
- *Concessions/licenses/franchises.* Although assets continue to be public and are "rented" to the private operator for use during the contract period, this operator can also bring its own assets. The concessionaire takes on operations and investment as well as commercial risk within the limits set in the contract, for a period that generally varies from 10 to 30 years. Subsidies can be part of the agreement, in particular when demand is not strong enough, implying that commercial risk is very high. Subsidies also can arise as a result of the contractual imposition of heavy service obligations. Because this is the most common form of contract, it is discussed later in this chapter.
- *Service contracts.* These are quite common in transport and deserve to be separated from concession/license/franchise contracts despite their strong contractual similarities. The main difference is one of scope and duration; both are smaller than for concessions/licenses. The government bids out the right to deliver a specific service and sometimes provides the assets needed. The winner can be made responsible only for costs. These are gross costs service contracts in which the government pays for the service rather than allowing the operator to collect revenue directly. The main disadvantage is that the provider is not interested in the demand for the sector, because it has the guarantee of public payment. This is why many governments prefer net costs contracts in which the winner is responsible for all revenue collection and costs (net cost service contracts). The main risk here is the temptation on the part of the winner to render the integration of a network difficult when it results in a more competitive provision of services.

means that there is now a choice of provider. Previously, governments argued that they were the only option, thereby holding users captive. It also means that the government's role in the sector must be revised. Because it clearly no longer has reason to be the single provider of most transport infrastructures, it has to focus on promoting competition among the various potential providers, rather than on micromanaging a strong monopoly as a regulator.

The new regulatory responsibilities are not that simple. Strong technological constraints still limit the opportunities for full competition in transport infrastructure. In the end, regulators must face the replacement of a "global" public monopoly by smaller, more specialized, private or even public monopolies. The limit to the size of this smaller monopoly is driven by the presence of strong indivisibility, joint production, and the difficulty of storing transport services.[2] This is essentially what drives the cost structure of transport infrastructures. These technological characteristics result in a joint cost component that makes specific tariff design quite challenging. Indeed, when some of the activities are delivered with the same inputs, the regulators have to be able to separate the costs of providing different goods and services to different classes of customers to identify undue tariff discrimination, cross-subsidies, and costs of universal obligations. This is because most of the required information is controlled by the operators, who often have little incentive to reveal it to the regulators. This is why these cost structures often are subject to arbitrary—and controversial—allocation rules across service lines, types, and user groups. Regulators face the challenge of minimizing the arbitrariness and the resulting inefficiencies of the allocation rules.

A final point is the importance of this challenge depending, of course, on the specific restructuring—the extent and ways in which competition is introduced in the sector—adopted for the sector and on the specific form of private sector participation adopted. Table 2.1 summarizes the major forms of private sector participation observed in the transport sector in developing and transition economies in this decade. Concessions are clearly

2. Indivisibility is the characteristic of an input into the production of a service that prevents its use below a certain minimum level. This is why firms tend to be large in sectors in which indivisibilities are common. Joint products are goods or services with the characteristic that a change in the rate of production of one brings about a similar change in the other; an increased number of terminals in a port requires an increased number of machinery and equipment to make the most of these terminals.

Table 2.1. *Types of Private Sector Involvement in Transport across Regions in Developing and Transition Economies*
(number of projects per contract types between 1990 and 1997)

Types	*Airport*	*Port*	*Rail*	*Roads*	*Total*
Divestiture	2	6	4	7	19
Greenfield	5	32	6	24	67
O&M projects	3	21	4	7	35
Concessions	15	31	23	170	239
Total	25	90	37	208	360

Source: World Bank PPI database.

the most common for infrastructure and services for private goods with strong demand. The remaining sector-specific chapters provide more details on the options and international experience.

The Diversity of Objectives Reflected in Economic Regulation

The main task of a regulator is to control prices and service quality and to make sure that the residual monopolists in the sector do not overcharge or cheat on the quality of service provided to the users. The regulators' role is to settle issues as specified in the charter or law creating the institution. In a nutshell, their decisions have to result in outcomes mimicking that of a competitive environment. This means that the regulator is concerned with efficiency and minimizing costs, while ensuring that investment decisions are consistent with demand at unbiased prices. In that process, a regulator must also ensure that the monopolistic operators get a reasonable return on their assets. Though it seems simple enough in practice, it is often more complex because governments face a multiplicity of objectives. Efficiency (stimulating cost minimization and pricing at cost) and fairness are only two of the objectives reforming governments seek. They also have strong fiscal and distribution concerns they want to address through the reforms. When this objective dominates the others, it can reduce the scope for efficiency, which can be quite constraining for a regulator.

To understand this better, consider the case of a port without any significant competition from other ports in its area of influence. The incentive of any operator would be to keep prices as high as the users are willing to go, generally much higher than the opportunity cost. When the price is much higher than the cost of providing services to a new user, the economy is wasting resources, that is, it is inefficient. The regulator can impose a price cap to avoid or minimize this waste. How high should this cap be? Clearly it should be higher than marginal cost so that it covers the cost of fixed assets. How much higher will tend to depend on the government's specific goals. This needs to be spelled out when privatization takes place to ensure that all potential operators know the rules of regulation. If the government wants the bids for the right to operate this monopolistic port to generate a huge fiscal revenue, it should probably allow a high residual monopoly power in the hands of the private operator. This power will allow the operator to enjoy a large wedge between cost and prices, which is what is going to make the operator bid a good deal to obtain access to this potentially large profit. The government wins, the operator wins (if the wedge between cost and price is not fully passed on to the government in the bid), but the users lose. The regulator cannot do much but enforce the commitments to a high cap that have been made by a government with a dominating fiscal concern.

Similarly, for firms delivering multiple services, setting the tariff in accordance with the opportunity cost will often imply different tariffs for different products. For instance, in the case of railways, the cost increases with the distance. For buses, the cost is inversely proportional to the traffic speed. These considerations imply that operators should be allowed to charge more to users in rural zones or in mountains where speed has to be much slower. This may not be politically desirable.To address the distributional concerns, the government may impose a mix of service-price requirements. This violates the efficiency concerns on which a regulator is supposed to be focusing. Most economists would argue that the government has better instruments to address these distributional concerns, but politicians often tend to adopt the solutions with the highest short-term payoffs. Imposing service obligations at the wrong price seems to be a common temptation that regulators will have to live with for some time.

More generally, these examples suggest that multiplicity of objectives is likely to be quite common and that regulators have to try to do the best they can under these constraints. Their job will be to make the consequences of these multiple objectives as transparent as possible. This will highlight the need for subsidies in some cases. In fact, some of the players will have to pay for the others when they impose inefficient decisions for fiscal or equity reasons. In some cases, this will show that short-term solutions hurt the incentive to invest, penalizing future generations of users in favor of the current generation.

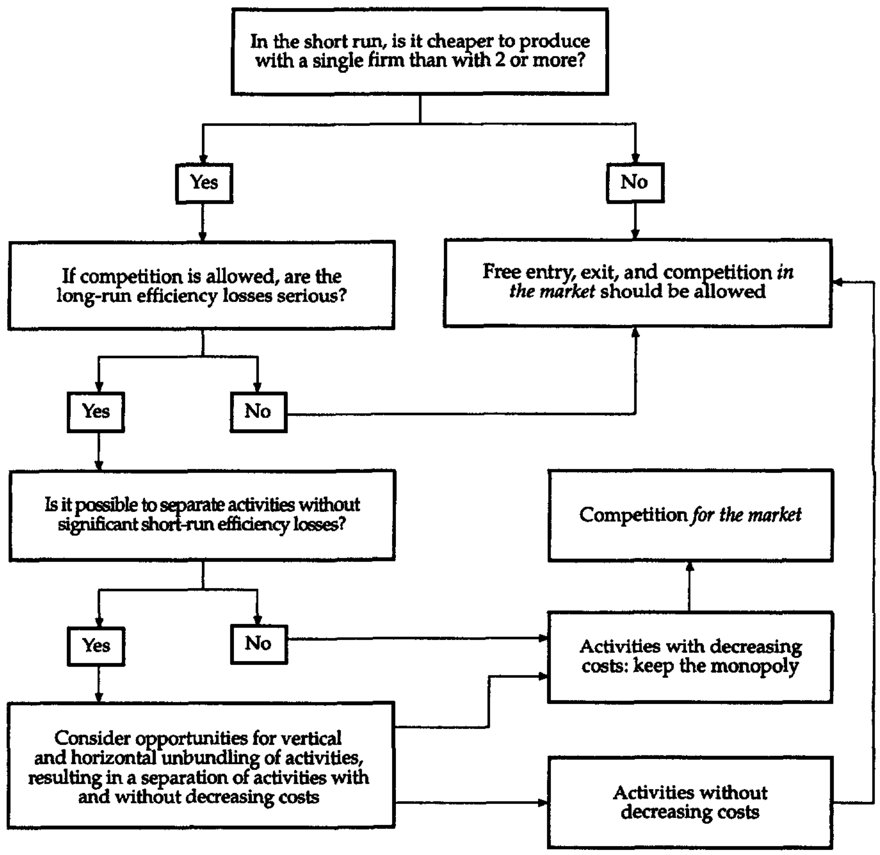
Natural Monopoly, Competition, and the Unbundling of Activities

As already hinted at, competitive markets can provide most transport services, and they are likely to require little or no economic regulation other than what falls under the realm of competition policy. Conceptually, the presence of decreasing costs associated with scale economies or indivisibility of assets, and hence the case for a monopoly and for its regulation, is almost exclusively associated with transport infrastructures. However, infrastructures are not monolithic. To assess regulatory needs and scope for competition, we distinguish between the fixed components—infrastructure in the strict sense—and superstructure. For instance, in a port, access roads to the ports are infrastructure while warehouses and mobile and immobile loading equipment are superstructure. Figure 2.1 shows the logic to follow to assess what type of competition is desirable, according to the degree of technical integration of the activity concerned and the desirability of allowing its provision by a monopoly. Whatever the response to the question asked, the ultimate outcome is to map the type of competition available to the technical characteristics of the activity. Indeed, if competition in the market cannot work—what most nonspecialists think of when they think of competition—competition for the market can be quite effective to obtain efficiency from a monopoly.

Competition in the market guarantees free entry and exit and lets demand and supply determine prices and quality mixes. Although introducing quality standards is often useful, this can be done without altering the nature of competition. This form of competition is effective in ensuring the long-run sustainability of efficiency gains. Competition for the market is organized through an auction used to force potential monopolists to compete with each other for the right to be the single provider of a service. The challenge is to design the auction to ensure that it forces the bidders to pass on many of the efficiency gains they should be able to achieve to the users, and to achieve results similar to those that would be achieved through competition in the market.

How does one decide how far to go in the restructuring of the sector to pick the right form of competition? It starts with a rather simple question: is it cheaper to produce with a single firm than with more firms? In theory, the question should apply to both the short run and the long run. In practice, the short-run concerns tend to dominate the longer-run needs of future generations. If two or more companies can do better than a single firm, interfering with competition in the market is not necessary. The only role of a regulator may be to ensure compliance with service obligations and government delivery on its promises to compensate the providers for these obligations.

Figure 2.1. *Natural Monopoly and Competition*



Source: Authors.

Even when leaving production to a single company garners efficiency gains, one may wonder whether these short-term gains are significant enough when compared with the longer-run gains that could be achieved through some degree of competition in the market. If these gains are not particularly impressive, one may try to push for competition to avoid having to deal with a monopoly, and to try to prepare the sector to reap the longer-run gains from competition. A first way to push for competition is to unbundle vertically and separate infrastructure and superstructure. This provides an opportunity to introduce competition in some specific activities. Horizontal unbundling also can help. For instance, this means separating passengers and freight in rail services. This separation between competitive and monopolistic activities is, of course, not that clean-cut. Some activities imply sunk costs,[3] but their economic importance is minor and generally insufficient to impede the entry of new firms in the market.

Competition still has a role to play once the activities have been sufficiently unbundled and the remaining natural monopolies have been identified—in other words, those bundles of activities for which the potential cost savings resulting from coordination and integration are significantly larger than the gains that could be achieved through competition. Competition for the market allows the regulator to try to get some up-front efficiency gains in the sector.

Unavoidable Restrictions to Entry and Regulation

Relying on competition for the market to assign responsibility for transport services not produced by a competitive market is an improvement over past practices, but it is also challenging, that is, the promise of improvement made by the private operator before it takes over the business. The main challenge from ensuring that the gains achieved ex ante through the auction mechanisms are maintained ex post (once this operator is actually in charge) stems from the exclusivity generally granted through the concession contract. The regulator must assess the specific implications of these conditions for whatever efficiency, equity, and fiscal goals the government may have. If too generous to the concessionaire, it may mean that the initial gains from competition for the market are probably not sustainable over the longer run. To a certain extent, the regulator and the government are captive to the concessionaires once the auction has been closed. Once a contract has been awarded, the concessionaire becomes the sole counterpart of the regulator. The concessionaire has a significant incentive to negotiate on anything that restricts its profits. This is one of the reasons why it is extremely important for the regulator to ensure that the concession contracts also include explicit rules for renegotiation of the terms agreed to when the concessionaire participated in the auction. It is also the main reason why it is important to make sure the government has identified a clear and valid reason for granting exclusivity, and to reject it if it is not needed.

---

3. Sunk costs are costs that cannot be recovered when an operator leaves the industry.

From the government's viewpoint, granting exclusivity may make sense for three main reasons as follows (Kerf and others 1998):

- *When cross-subsidies among different users are denied.* This happens in three types of situations involving cross-subsidies.[4] First, certain user groups may be required to have tariffs lower than costs, and other user groups may then be allowed to have higher tariffs to compensate. Second, a concessionaire may be required to have a unified tariff for all users, despite differences in costs to serve across user groups. Third, existing users may be asked to subsidize the expansion costs needed to have more users in the network. In each of these cases, the exclusivity avoids unfair competition from firms able to focus only on new users without the burden of having to deal with the cost constraints imposed by the contract on the concessionaire. The exclusivity guarantee makes cross-subsidies viable for governments unable to assist unfavored user groups through direct subsidies.
- *When the initial risk levels in the sector or the country are high.* Because competition tends to reduce benefits, exclusivity makes a concession more attractive in an auction, which can improve competition for the market because more bidders are likely to show interest in the auction. In many developing countries, temporary exclusivity conditions are sometimes the only way to ensure participation in an auction.
- *When the service the concessionaire will provide is a natural monopoly.* Firms are sometimes interested, for strategic reasons, in entering a market in which, technically, there is room for only one firm. This is a case when the government should rely on exclusivity to avoid undesirable entry into the business.

Once more, these are valid restrictions on competition that governments may decide to adopt. A regulator cannot do much but take them into account. A major responsibility these specific restrictions impose on the regulator is the measurement of the importance of the cross-subsidies allowed. This, in turn, requires close monitoring of the costs of service of the various user groups, which is one of the toughest challenges a regulator will face, because the providers have control over the cost information per user group. The privatization design needs to anticipate this to require the monopolistic

---

4. This is not an endorsement of cross-subsidies. It is simply a recognition that cross-subsidies sometimes arise when governments cannot rely on targeted subsidies for fiscal reasons.

operators to reveal enough cost information to allow the regulator to minimize the misallocation of resources that can result from the exclusivity.

The Main Elements of Contract Design from a Regulator's Viewpoint

While writing a contract that is "all encompassing" is often difficult, every contract must address a minimum set of issues if this contract is to be a useful tool for directing and supporting regulatory decisions. This section summarizes the minimum requirements.

Basic Coverage of a Contract

The contract is the legal instrument spelling out the key economic elements that the government wants to cover in its agreement with the private operator. The coverage of the contract must then be all-encompassing and include a detailed description of the object of the auction, the obligations and rights associated with this object, the processes to follow, and contingencies in case of unforeseen events. All parties involved—privatization commission, investors, operators, users, and taxpayers—should have clear legal support for the economic and financial agreement. This is particularly important in countries in which the legal system is not oriented toward contract law. Regulators need to be able to rely on the contract to shortcut excessively complex legal systems designed when governments wanted to support the public operation of infrastructure services and exclude all private roles in these sectors. Contracts also need to provide a set of instruments for the new regulators of the sector that allow fair and efficient regulation within the constraints explained earlier. From a regulator's viewpoint, a contract must cover the elements discussed in the following paragraphs.[5]

Description of the Specific Activity Covered by the Contractual Agreement. This is the coverage of the services and the size of the market, which is much more subtle than it first seems. It needs to be specific and answer such questions as, "Does the contract cover service/traffic lines or geographic

5. See, for instance, Crampes and Estache (1998); Gwilliam (1998); Kerf and others (1998); and Shaw, Gwilliam, and Thompson 1996. For a more detailed overview of what contract design is in practice, with an illustration of water and sanitation contracts, see Brooke and others (1997).

zones?" and, "Is the contract covering all zones to be contracted or only parts of these zones?" The first question tells the regulator whether it will have to take on the coordination of the physical network (the timetable for the use of rail tracks or for bus services, for example) while promoting the development of a network of operators capable of competing with each other and providing a spectrum of services. Will it be the responsibility of the single operator for all lines included in a region? The second question also tells the regulator whether it will be able to easily compare the performance of activities in various regions (road or port concessions in different provinces) or lines (various intercity train or bus services) or whether it will have to interact with a single monopolistic provider of all services. In a nutshell, this first item in a contract is the outcome of the form and extent of horizontal and vertical unbundling that the privatization team has adopted, and it defines the limits and scope of the regulatory activities.

EXCLUSIVITY. As mentioned earlier, what can make or break competition for the market is the degree of exclusivity over the right to provide a specific service (for example,track or road maintenance, among all the services required from the contractor for the operation of a train service or road concession) and not necessarily all services granted by the contract for a specific period of time. The contract draft passed on to potential bidders must be quite clear on the specific activities, services, or geographical zones for which there is exclusivity, both to ensure the effectiveness of this form of competition and to ensure that the regulator has clear terms of references over its assessments of the rights and obligations of the contracted firms. It is also a key instrument for a competition agency that has been asked to assess the legality of restriction to entry in the sector introduced by the incumbent. In general, the shorter the exclusivity period, the lower the risk of conflicts. When the exclusivity period is for the duration of the contract, and the contract is long, the risks are much more serious of having frustrated potential entrants excluded from a market in which consumers would benefit from entry as a result of cost reductions.

ASSET OWNERSHIP AND VALUATION. Being able to assess the assets is quite important for the privatization team, because it sets the minimum price the government is willing to accept for the activity it is contracting. But it is also important from the regulator's viewpoint and matters enormously at the end of a contract or during conflicts. If the assets are public and little new investment has occurred during the contract period, the regulator is only accountable to the government when controlling that the state of the assets returned to the government is consistent with the contractual demands. This in itself is a challenging task, because private operators generally have little incentive to spend money to properly maintain assets they will soon no longer need. When investment takes place and/or the private operator brings in assets, the regulator also becomes accountable to this operator to ensure that it is properly compensated for all investments or assets not yet amortized through tariffs. This means that the asset valuation methods adopted have to be quite clear to all parties involved, which is a complex and controversial matter (see Burns and Estache 1998).

In many countries a particularly troublesome component of this process has been the assessment of demand. The incentive of privatization teams is often to overstate demand to increase the value of the business. This is why regulators in so many countries are faced with contract renegotiation requests because the private operators end up finding out the hard way—on the job—that the user's willingness to pay is not as high as initially anticipated. To remedy this risk, privatization processes are now increasingly relying on specialized consulting firms to carry out the demand studies. In addition, many of these firms are now starting to sell insurance to cover the demand risks. This is a significant improvement from a regulatory viewpoint, because a key source of conflict should disappear with the creation of a market for insurance.

Duration. As a rule of thumb, the shorter the duration of the contract, the stronger the potential for competition for the market, because the activities object of the contract will be subject to more frequent auctions. This is particularly true for activities involving few sunk costs and little asset specificity, and for those facing a lot of uncertainty. Urban bus services are a good example, because buses can be easily relocated to other cities or used for other types of services. So why are all two-to-three-year contracts not subject to frequent auctions? The main reason is the difficulty of convincing the winners of short-term contracts to make investments for the long term. The ideal contract duration is long enough to allow the amortization of investments made and a fair return on the investment for the given pricing rules spelled out in the contract. This will remain a major concern as long as asset valuation remains a source of conflict between regulators and operators, and as long as investors are not convinced they will get a fair compensation for investments made, but not amortized, at the end of the contract.[6]

---

6. One cannot discard the concern for political risks and fear of political interference with the implementation of the contractual commitments that are more likely to arise with long contracts, because they expose investors to more political cycles.

This concern is valid whether the government is responsible for compensation at the end of the contract or whether the winner of the next auction for the contract is responsible. Moreover, in addition to the obvious high transaction costs associated with frequent auctions that have to be balanced against the gains from competition, short contracts can be a major source of concern when demand is uncertain. The Mexican toll road experience has become the standard example to illustrate that well-intended short-term contracts can fail when demand is very sensitive to toll levels. Because short contracts with strong unsubsidized investment obligations will generally mean high tolls (needed to recover the investment during the contract period), a careful assessment of whether the willingness to pay is high enough is important. In many cases, demand will not follow and contract duration will have to be adjusted through a renegotiation.[7]

This is why one solution for certain types of activities is to adopt variable-duration contracts. The difficulty of forecasting demand with long lead times (and hence of estimating long-term revenue) and of anticipating demand shocks that are uncontrollable by the operator (recessions, financial crisis) is addressed by a methodology Engel and others (1996) proposed for the road sector. Various Latin American countries have adopted it, starting with Chile. The idea is to have the regulator set the toll and the discount rate, in addition to all the service and investment obligations, and to have the potential operator compete in an auction in which the winner demands the lowest net present value for the future revenue to be collected through the contract. The duration becomes variable because the operator returns the road to the government only when it has cashed in the revenue bid. When demand is strong, the contract can be short. When it is not, it can be long.

Moreover, if for any reason the government wants to change anything (such as expanding the road), one can easily assess the financial consequences in terms of the net present value. The outcome is an automatic adjustment in the contract duration. This flexibility has reduced the risk premium demanded by investors, and hence the required rate of return. This arrangement, which works well for simple activities such as roads or

7. Remember that the privatization process generally leads to a situation in which users end up paying for a service for which they were either not used to paying (roads) or for which they were hardly paying anything (trains and buses, which tended to be paid for by taxpayers). This often leads to strong user reactions that initially can curtail demand if the operator cannot tailor the service price mix well as a result of excessive contractual requirements for the contract period allowed.

airport runways, is not foolproof. Its main problem is that because revenue is guaranteed, the operator does not have much incentive to maintain the infrastructure quality. That is why in these types of contracts, quality requirements (essentially maintenance) and the related penalties for noncompliance are important.

INVESTMENT AND OTHER OBLIGATIONS. The contract can also be used for specific investment projects or increases in capacity that the winner needs to provide. New investments often have a strong temptation to be quite specific about all the technical parameters of the investment. This often results in the regulator micromanaging the operator's investment decisions. The trend in government bias is generally to push for overinvestment in quality, inconsistent with demand. More specifically, there is always a risk that this overinvestment will result in tariffs that the users are not willing to pay. The general advice given to governments is to focus on outputs, in terms of service coverage and quality, and not on inputs. However, the regulator wants some guidance when monitoring compliance with these obligations. General targets are useful in this regard. This is particularly important at the end of the contract when the regulator must ensure that investments have been properly maintained. Without specification of what proper maintenance should be, the odds of conflict with the operator are high.

A common criterion is to rely on third-party assessment of technical quality. One can also use this third party to assess when cases of *force majeure* justify delays in construction or delivery of service commitments. Finally, a particularly common event to avoid, one that has gotten regulators in tough situations, arises when investments and service obligations are contingent on financing possibilities. In risky situations, bidders should have financing for their investments lined up. When the government shows unpredictability in handling a privatization program or when the global environment becomes too chaotic, this financing may not materialize. Often it will also be reasonable for the government to minimize the risks of having to organize a new auction and specify in the contract appropriate contingencies when this risk is serious. The government may also want to specify performance guarantees in the contract to provide the regulator with a clear instrument to penalize in cases of nonperformance not covered by the contingencies.

REVENUE, TARIFFS, AND REGULATORY REGIME. This is probably the most complex element of the contract from an economic perspective. It covers both generic and technical questions that influence the effectiveness of the contract. Prices are clearly driving revenue. The main point to address is often a technical "pricing" annex attached to most contracts, because this annex is one of the main instruments of any regulator. In addition to the definition of the pricing rules, including indexing and the decision to control average versus individual tariffs, investors may need to get clear guidance about some basic concerns to assess their revenue prospects and flexibility. The first concern to clarify is whether they can be paid offshore. This has multiple implications from the investor's viewpoint, not the least of which are tax liabilities. It is also a difficult issue for the host governments because they do not want to risk conflict with the country of the investors. At the same time, this can be an effective risk mitigation strategy resulting in a lower required rate of return by the investors. The regulator will have to account for this input in calculating the rate of return.

A related concern that needs to be addressed early on is currency: which accounts must be held, are the tolls in dollars or in local currency? Finally, this annex must clarify the technicalities: it must spell out the regulatory regime (price cap or rate of return, or a hybrid?) and the extent to which price differentiation is allowed or imposed (for example, to meet social concerns). These more technical questions on the choice of the regulatory regime, and on the extent to which tariff design can and should be covered by the contract, are discussed in detail later.

CONTROL AND SANCTIONS. The discussion so far has made it clear that the operators and regulators face many uncertainties, because contracts will often not be able to cover all contingencies. Moreover, the regulators will not always be able to control everything on a continuing basis. Operators know much more about the business than regulators do and excessive control would be equivalent to micromanaging the business. The operator must be left to work on delivering its contractual commitments at a reasonable, regulatory compliance cost. Some degree of control is needed, but it will only be effective if the regulator has a clear set of sanctions to apply. For example, performance bonds or guarantees are part of the standard kit of sanctions for operators who do not deliver on time. But what happens when assets are not maintained properly or when service quality (such as unacceptable delays) or safety standards are not met? The ideal sanction is one that compensates the victims of the operator's failure to deliver on its obligations for the equivalent of the loss incurred by the victim. This raises complex methodological issues as well, but it can be reasonably approximated for most minor contractual violations.

In general, sanctions cannot be so low that the operators ignore them and treat them just as built-in operational costs. Similarly, they cannot be so high that they force the operators to close shop as soon as they do not comply with the contract. Note that to ensure the fairness and transparency of the process, this section of the contract must also be clear about the public hearing and appeal process when sanctions have to be decided. This is important, because in transport, the nonrespect of obligations is often not necessarily the operator's fault. For instance, unexpected traffic can cause delays due to public works that were not expected at the time the contract was signed and that could not be built into the original timetable. A solution to stimulate operator creativity in addressing these unexpected events is to combine incentives for meeting quality indicators with sanctions for the most obvious deviations from commitments.

RENEGOTIATION. This happens quite frequently, and the contract should clearly specify its terms. It often occurs because demand is much lower than expected and the operator wants to slow down the investment program. This has been the experience for many toll roads. However, demand may also be higher than expected and investments have to be accelerated and/or tariffs increased. These cases are relatively easy to assess. Often, however, the regulator has to make a fair assessment of the need to renegotiate, assess the costs and benefits, and determine the winners and losers. Governments commonly ask for faster investments and lower tariffs as an election comes closer. In that case, a fair regulator will often have to ask the government to compensate the operator for the financial consequences of its renegotiation request. As a rule, the party asking for a renegotiation will have to compensate the other for the consequences of its demand. As a second rule, the renegotiation should generally not change the net present value of the business for the investor. As is becoming increasingly clear, the legitimacy of the privatization processes depends, to a large extent, on the way in which governments leave regulators to handle the renegotiation, and on the way in which these regulators perform. What seems clear is that allowing renegotiation under broad or vaguely defined circumstances makes a mockery of the competitive bidding for concessions.

TERMINATION CONDITIONS. A final, basic aspect of a transport contract is the definition of the termination conditions. Is the renewal of the concession or service contract automatic? Is this negotiable? If it is negotiable, what will be expected from the regulator in terms of asset or business valuation? Increasingly, however, privatization commissions are recommending that new auctions be organized to select the new contractor. Under any renewal strategy, the contract must be clear on the conditions of transfer of assets. This means specifying the expected state of the assets, and whether they will be returned to the state or to the new contractor. Finally, this part of the contract must also address the possibility of an unanticipated end to the contract. The contract must recognize that the unanticipated termination may be due to either the government, the contractor, or to a mutual agreement. Under any scenario, whoever initiated the early termination will be required to make compensation. Ideally, these payments should be formula-driven, and they should consider the associated residual asset value and transaction costs.

Criteria for Organizing an Auction and Picking a Winner

Organizing an auction and picking a winner can be done in many ways. Choosing the winner at an auction for a concession or a service contract requires a good understanding of the trade-off involved. In practice, however, in countries in which the governance structure is not reliable or well tested, the rule of thumb to follow is straightforward: keep it simple, fair, and transparent to maximize the number of bidders, ensure the success of competition for the market, and minimize the risks of corruption or unfair decisions.

The process is generally as follows. The government provides information on the state of the assets, the value of the business, and its contractor requirements to identify the potential bidders. For large projects, this can take the form of a road show to allow a prequalification procedure, which is essentially a marketing trip to the regions where most potential investors are located, often France, Germany, Japan, Spain, the United States, and the United Kingdom. This stage also includes criteria for a technical and financial prequalification of the bidders. The potential bidders are told the type of sanctions the regulators may impose if they do not deliver once they have won the contract. This prequalification is not always necessary, but it is recommended. It minimizes the risks of having incompetent, risk-taking investors trying to get into a business they do not know, or of competent operators trying to commit to financing they cannot deliver. In instances in which this stage had not been taken seriously, governments have been embarrassed, having to declare the auction invalid after finding out that the winner could not deliver and was just trying to get a contract that it hoped to immediately renegotiate. Generally, the prequalification criteria result in the creation of consortia, pulling together investors and operators.

The next stage is the definition of the auction itself. It is organized around two criteria: technical and financial. These can be assessed in three ways as follows:

- A two-stage selection that first eliminates the weakest technical proposals, then picks the best economic offer among the remaining candidates.
- A single-stage proposal that ranks the offers according to a weighted average of the technical and financial proposals. This is often very tricky and can result in unfair decisions when there is a choice between widely differing technologies with different financial consequences.
- A procedure in which the government specifies the technology and all the engineering aspects desired, and the bidders only compete on the financial dimension.

To actually pick a winner, more specific criteria are needed. This depends on the specific objectives the government is trying to achieve. Table 2.2 suggests the optimal selection criteria for a spectrum of possible objectives. While several of the objectives are closely related, they are separated because they reflect different terminology politicians use when justifying the privatization strategy adopted.

The table shows that trade-offs do indeed exist. For instance users are not guaranteed to benefit from lower tariffs if the government has fiscal objectives in mind. Indeed, the best way for a government to maximize the willingness of bidders to pay in an auction is to offer a strong monopoly, which implies large potential profits through high tariffs. This strategy makes it particularly difficult for a regulator to ensure efficiency and fairness because the contractor gets contractual protection for the right to use its monopoly powers. Governments often justify this as a temporary negative point in return for the long-term good of the nation. This ensures that the private sector will make the investments that the public sector could not finance. In cases of conflict, however, this is often a sore point, and it threatens the credibility of the privatization program. At the other extreme, the government wants to minimize political conflicts with the unions, for instance, by imposing the retention of unneeded workers simply to increase the costs of the service. Users continue to pay for the failure of the local labor market to absorb excess employees.

Remember that multiple award criteria usually do not work. They require arbitrary weights for the various criteria, which tends to lead to arbitrary selections. Any arbitrary selection criterion will introduce a risk of corruption. In sum, the individual circumstances of a country or a sector

Table 2.2. Relation between Privatization Objectives and Auction Award Criteria

Privatization objectives	*Auction award criteria:* *Shortest contract duration*	*Lowest tariff*	*Minimum subsidy requested*	*Maximum payment to the government*	*Maximum number of employees retained*	*Best investment plan*	*Minimum net present value of revenue requested*
Competition	•	•	•	•			•
Infrastructure quality and capacity						•	
Benefits to the users		•				•	•
Reduction in fiscal deficit			•	•			•
Minimal political conflict					•		

• Indicates that the linkage is close.
Source: Authors.

should clearly determine the choice of the specific award criterion. The net present value criterion offers many advantages when awarding contracts for the operation, maintenance, and development of simple infrastructure such as roads, runways, or ports.

Risk Assignment

An important component of the information transmitted to the potential bidders through the draft contract is the allocation of risks among the government, the operators, and the users of the service. This is a complex matter. This brief section can hardly do justice to its importance but will highlight it from a regulator's viewpoint. Indeed, identifying the various types of risks, and their distribution across the various agents, is important because it influences the incentives these various agents will have to behave in one way or another on regulatory matters. For instance, if a concessionaire is allowed to pass through all increases in costs because of changes in safety legislation, it will have little incentive to pick the most cost-effective technologies, because it does not bear the costs of its choice. This is why the British airport regulator only allowed a pass-through of 95 percent of these costs when a new safety norm was introduced for airports in 1996. The competition agency's initial recommendation was actually 85 percent.

A full pass-through makes sense only for failure to comply with the contract resulting from risks completely out of the concessionaire's control, such as floods, national strikes, or earthquakes. Otherwise, the risks should be assigned to the agents most capable of their assessment, their control, and their management so that their cost is minimized (Kerf and others 1998). Depending on the specific risk and its source, some standard recommendations are available for risk assignments in concession contracts (World Bank 1997). They are summarized here because regulators may have to assess the nature of risk and assign its responsibility in cases of conflict.

- During the design stage, specification failures do arise in the bidding documents that the government provides, and they are clearly the responsibility of only the government. If the failure is in the design proposal as part of the bids provided by the bidder, it is the bidder's fault.
- During the construction stage, legal changes or difficulties and delays in expropriations of land can increase the costs of a project, but the contractor is generally not responsible. The government or an

insurance company should thus cover this risk. Construction difficulties caused by technical failures in the choice of material or equipment are, of course, the constructor's responsibility.

- During the operational stage, cost risks should be the the operator's responsibility unless cost overruns are due to the government's failure to deliver on a specific commitment (such as delivering a permit or providing inputs promised on time) that results in costs increases. Revenue risks should also be the operator's responsibility unless the contract specifies otherwise,[8] or when the failure to generate the expected revenue is the result of a government action (such as failure to increase tariffs according to formulas specified in the contract).
- The contract should also specify the responsibility for financial risk and exchange rate risk, which is often a subject of negotiation between potential bidders and the government. Only in cases of the introduction of convertibility restrictions is the government clearly responsible for indemnizations. For the rest, governments have often been too willing to cover these risks, promoting careless behavior on the operators' part as a result.
- The responsibility for environmental risks should rest on the concessionaires or their insurance companies (as many of the risks of *force majeure* that insurance companies are willing to take on).
- Expropriation risks should generally be the government's responsibility. The challenge lies in the enforceability of this risk allocation.

Price and Quality Regulation

The main reason that direct competition between potential transport service providers is increasingly viewed as desirable is that the freedom these providers have to set prices tends to benefit users (who can pick among a wider range of price-quality mix). In addition, it ensures that the providers will have an incentive to minimize costs while setting prices that guarantee their financial equilibrium. All this can be achieved without a regulatory authority drastically controlling prices in a competitive environment. If an airline or a trucking company does not generate enough revenue because it is not competitive, it goes bankrupt. This

8. Which may be the case when governments perceive that without some sharing, the private sector will not be interested at all.

occurs because its costs are too high, demand is weak, or technological changes have taken place.[9]

When competition in the market is not possible and competition for the market results in a legal private monopoly, as is the case with most concession programs, regulating the prices that this monopoly will be allowed to charge is one of the mechanisms that make the defense of the users' interest consistent with the tolerance of a private monopoly, including one with exclusivity rights over a specific market. If the regulator had all the necessary information, it could ensure that the users would get an outcome close to the one that would emerge from effective competition by allowing prices to cover the average cost. Unfortunately, the regulators do not have access to the same details of information about technology, cost structure, and demand that the airport, port, and road operators do. This is why one of the main tasks of a regulator is getting the operator to reveal enough information to demonstrate that it is not abusing its power over the market.

Financial Equilibrium and Prices

Consider the case of the most common regulator. What is it supposed to do when the contract does not specify prices precisely enough? It is trying to get the operator to minimize costs and allocate resources to where they yield the most to society (static efficiency in economic jargon), and at the same time trying to stimulate the right amount of innovation and investment to ensure that the operator can meet future demand (dynamic efficiency). This regulator knows that it has to allow the operator to at least break even, in other words, guarantee cost recovery. It also knows that it is likely to have to meet some social concerns and cannot end up micromanaging the operator and imposing excessive regulatory compliance costs.

The challenge comes from the fact that the monopolistic operator's interest is not quite the same. The monopoly wants to maximize profits. The prices it will prefer to charge will be compatible with cost minimization and cost recovery. It will be higher than what is needed to achieve an efficient

9. The practice of competition is, of course, more complex, and many perverse forms of competition can arise from some firms strategically trying to exclude potential entrants from their markets. Serious risks of collusive behavior between potential competitors also exist that could offset the potential gains from breaking a monopoly. This is why governments have a major role in monitoring the behavior of competitors through an effective competition agency.

allocation of resources, to meet demand, to promote investment decisions consistent with demand, or to meet social concerns. In general, the regulator will have to introduce a system that reduces prices or changes the price structure to reconcile the monopolist's right to a reasonable rate of return on its investments with the interests of society.

To ensure the regulated operator's financial equilibrium, the total allowed revenue must be at least equal to its total costs. A disaggregation of the components of revenue and costs most relevant to the regulator suggests the following simplified formula:

(2.1) *price x quantity = operational costs + (asset value x the cost of capital)*

This equation shows that the cost structure of the operator has two main components: operational costs and capital costs. The first set of costs includes all the standard inputs and the volume of production generally allocated to each activity, for instance, the cost of bus drivers can be assigned to each line. They can also be common costs, which are much more difficult to allocate across activities, for instance, administration costs for the management of this bus company.[10] The variable costs are usually well handled by standard accounting procedures. Common costs are generally handled in a much less satisfactory way by financial accounting. For regulatory decisions, regulators need to strongly guide their accounting.

The second set of costs is capital costs. These are even more complex to deal with from a regulatory viewpoint because they are much more difficult to assess. First, the value of the assets must be estimated. A base value is usually set during the privatization process, but it is often subject to revision once the private sector has taken over the business. Next, the regulator must provide guidance on how to calculate the costs of capital—essentially the minimum rate of return that makes it worthwhile for the firm to stay in business. Both values are subject to serious controversies.

The main problem in assessing the value of the assets stems from the fact that various methodologies exist, each with clear biases, that are hard to quantify: historical value, market value, and replacement value. In addition, the privatization commission often tends to overstate the value of the assets when it passes on the business to the private sector to increase the fiscal payoffs. The operator has a similar incentive because a high initial value leads to high recoverable costs and reduces the risks of a

---

10. Common costs are costs that are common to various business lines of the company.

high capital gains tax when it resells the assets. The losers in this case are the users and the taxpayers. The regulator is the only champion they have to defend their interests. This defense takes place every time a tariff revision occurs, because it provides an opportunity for a transparent reassessment of the assets' value.

The cost of capital can be set in relation to the opportunity cost of the resources invested, but no single value to assess this opportunity cost is available, because it is project-specific and depends on the risk allocation among the operator, the users, and the government. The lower the risk sharing, the higher the variability of revenue and costs expected, the higher the costs of capital will be. In general, the regulator is going to have a double, long-term objective: the first to ensure that the operator gets a reasonable rate of return on investment, the second to make sure that this return is not excessive. The first objective is needed to ensure that investment actually takes place. Indeed, without a rate of return or, equally, a cost of capital sufficiently high, no investment will take place. The second objective is to ensure that the operator with monopoly power does not abuse this power. These objectives can be achieved in many ways. The two extremes are rate of return regulation and price caps. The other options tend to be hybrid solutions.

Rate of Return Regulation

Until recently, the main approach to monopoly regulation was the control of maximum rate of return allowed from investment. It is essentially an indirect way of controlling prices, because prices above the competitive prices will result in an above-normal rate of return of the sector. The allowed rate of return determines the allowed profits of the firm, as illustrated by equation 2.2:

(2.2) *allowed rate of return x assets value = prices x quantities – operational costs*

This expression implies that a firm will not be interested in the business unless

(2.3) *prices x quantities ≥ operational costs + allowed rate of return x assets value.*

Both expressions illustrate the difficulties of the regulatory process. They show that rate of return regulation requires, first, detailed information on costs. Once operational costs have been assessed, the regulator has to assess both the assets value and the cost of capital to assess the minimum profit compatible with private investment. Because the objective is to allow a normal rate of return, which is assessed ex ante, the revenue must at least cover total costs. This is quite attractive for investors because the regulator will generally allow a risk-adjusted rate of return, implying that the cost of risk is passed on to users. Indeed, given a forecast of the volume of traffic, the price allowed is determined as a residual to ensure that the equation holds.

The main problem with this indirect form of price regulation comes from perverse incentives built into equations 2.2 and 2.3. The larger the value of the asset, the larger the benefits allowed, and hence the higher the prices will be. This can result in an incentive to overinvest (the Averch-Johnson effect), or simply to overstate the value of the assets when their correct value is difficult to assess precisely. In addition, equation 2.3 shows that the operator does not have much incentive to cut costs, because the larger the costs, the larger the benefits allowed. The system penalizes efforts to cut costs, because they would have to be passed through to users immediately through price reductions. A final drawback is excessive compliance costs. Every year the regulator demands detailed information on costs, assets, and investments to assess the required price adjustment.

These problems have led some regulators to adopt adjusted versions of the rate of return approach, allowing the operator to share in the profits resulting from cost or price reduction, generating a significant increase in demand.

Price Cap Regulation

The United Kingdom introduced an alternative to rate of return regulation as part of the privatization of the 1970s and 1980s in various sectors, and it is now becoming common worldwide. It increases the cost incentives and reduces the incentive to overinvest. It is based on the control of maximum prices or, in economic jargon, the imposition of price caps. In a nutshell, a price cap allows an operator to increase its prices with inflation, less a "discount" reflecting all or part of the average increase in productivity (a factor X) in the sector. This factor is introduced to ensure that the gains from technological improvements are not simply an increase in the monopoly's profit, but they also benefit the users. In the case of industries with little capital, X can be negative, allowing increases in real prices intended to stimulate new investments or to improve quality of service as imposed by the contract.

To present the conceptual elements of what the price cap is all about, equation 2.4 describes the basic principle for a monopolist providing a single product:

(2.4) *price in year 1 ≤ price in year 0 x (inflation–factor X)*

The first challenge is to define price in year 0. If prices were controlled and dramatically subsidized before privatization, the regulator cannot rely on current observed prices because they are probably too low. The regulator must be careful not to be too strict on the cap. If the firm cannot cover all costs, including the risk premium and the real costs associated with bankrupcy, it will enter into renegotiations, which are always difficult and may lead to suspension of maintenance and investment. This can be damaging to the network. When the initial conditions are too confusing, relying on best international practice as a first step is often a good option.

Next, the regulator must pick X. When X is set to 0, prices are simply adjusted for inflation, maintaining the real price as a constant. When efficiency gains are expected in the regulated industry, a positive X will lower real prices. This should stimulate the firm to cut costs, and to achieve efficiency gains higher than the industry average. A negative X will increase real prices and the regulator should use it when it wants to promote additional investment in capacity or quality.

Once the regulator has fixed X, it is usually kept constant for four to six years and the regulator does not adjust prices to reflect efficiency gains for that duration. This stimulates firms to introduce efficiency improvements and results in cost reductions in that period as soon as possible, because it will increase the new present value of its profits. After that period, the regulator revises the X based on observed cost reductions and passes on all or part of these gains to the user by resetting the value of X appropriately.

For a given X, the longer the time spent between setting the initial price and setting the price revision to redistribute the efficiency gains to users, the larger the incentive to cut costs. Cutting the less sensitive costs is to the firm's benefit. The stronger the demand elasticity for the activity regulated, the larger the increase in the use of a transport service for a given reduction in price, the shorter the period should be to reduce the risks to which the firm is exposed. However, this increases the costs of regulation. High compliance costs are one of the criticisms of rate of return regulation. A longer revision period can be obtained by allowing the firm to pass through to prices some of the costs not under its control—fuel costs or costs imposed by new legal environmental requirements, for instance—and that are less sensitive to efforts to cut them. Note that long spreads between revisions are sometimes difficult to handle politically, in particular when profits appear to be very high. The public wants to see some of them passed through to users through lower tariffs as soon as possible. Box 2.2 shows how this is done in practice.

> ***Box 2.2.*** *Price Caps in Practice*
>
> A useful illustration of these principles is the control of tariffs charged to access the rail owned by the Railtrack company in England. The rail regulator announced in 1995 that it would allow the tariff to increase to the maximum, with the inflation measured by the consumer price index less 2 percent every year between 1996–97 and 2000–01. The regulator also allows increases in tariffs reflecting increases in costs due to legal obligations imposed on Railtrack and for changes in energy costs up to the average price increase that industrial users paid for electricity. The access charges would be set to cover operational costs, amortization costs, and the return to capital. Once the value of Railtrack's assets was established, the initially allowed rate of return was set to 5.11 percent, hoping that it would increase to 8 percent in three years. This was compatible with a normal rate of return for this type of investment. The regulator also imposes that extraordinary profits be shared with the clients through reductions in access charges. Imposing explicit minimum service indicators (time of travel and quality of the ride) decreased the risk that Railtrack would try to cut average costs by reducing quality.

Consider now the most common case of a monopolistic concessionaire or contractor enjoying the responsibility for multiple services. Here, instead of using the price of the single product as above, the regulator sets the cap for a basket representing the bundle of services provided by the operator. The price of this basket is indexed to inflation—most commonly the consumer price index—less the minimum efficiency gains the firm will have to achieve across products if it does not want to see its average real price decline. In practice, the regulator first sets the cap in year 1 for the bundle of services provided by aggregating the prices observed in year 0 with weights reflecting the relative importance, generally in terms of sales volume, but any other reasonable weight can be picked. This can be changed at the tariff review of each specific service in the operator's total business. The regulator then announces the efficiency gains that will be assumed to have been achieved for the period, which will be passed on to users at the next tariff review. Neither the service bundle nor the weights of X can be changed in between two tariff revisions. This commitment is crucial to the credibility of the system, because it provides guarantees that the operator must try to make all possible efforts to cut costs quickly.

A useful quality of regulating the average price, rather than individual prices, is that it allows the firm to decide on its own price structure and to adjust better to price demand sensitivities. This can be problematic, however, in cases in which social concerns are important. The poor, with few alternatives to using services provided by a monopolist, may end up paying prices that are much higher than costs, which is why regulators sometimes leave socially sensitive services out of the basket and set individual caps.

Note that when a large proportion of the costs do not vary with the volume of service provided, regulating the maximum revenue instead of the maximum price is sometimes useful, such as for the regulation of airports. The main problem with a revenue cap is that it can reduce the incentive to maximize sales and can give concessionaires opportunities to play with the price, quality, and volume mix in undesirable ways. A contractor will often prefer to sell more of the services with below-average prices, low marginal cost, or profit margins, rather than high costs and high prices with lower profit margins.

Regulating Service Quality

Quality is a complex concept. It involves safety, ease, reliability and type of access, and the type of interactions between operators and users.[11] It also involves environmental concerns. How much a regulator needs to take into account these various facets of quality depends on many factors, including the degree of competition in the industry for most indicators.

SERVICES DELIVERY WITH COMPETITION. When the services are competitive, unless some monitoring takes place, private operators can be unsafe, unreliable, or simply unpleasant in their interactions with customers. In Sri Lanka, for instance, competition for passengers among small, private bus operators serving the same route has resulted in drivers racing down the road to pass a rival and beat it to stops or a timetable that results in increases in accidents and fatalities (Gomez-Ibanez 1997). Under better-monitored competitive pressure, users can also get service providers to reveal enough information on quality. When airlines compete for customers, they are happy to detail favorable performance comparisons, and users get to know more about the services they are buying. These suppliers, however, still control and sometimes restrict access to information on the health, safety, or environmental implications of the way in which they deliver their services. Therefore, regulators should address these aspects of service quality in an explicit way. Often, this requires coordination with other governmental agencies, as discussed later.

SERVICES DELIVERY BY A MONOPOLY. When a monopoly delivers the service, the challenge is more complex. The level of service may vary according to

11. For a longer discussion, see Carbajo, Estache, and Kennedy (1997).

how dependent the user is on that service, as well as on the existence and design of regulation. If left unregulated, a monopolist is unlikely to provide the exact level of quality the users demand. Whether this unregulated monopolist over- or underprovides quality depends on users' responsiveness to changes in prices reflecting changes in the quality of service delivered. This is an empirical matter. In general, however, the monopolist will have to be regulated. How this monopolist can be and should be regulated is not as straightforward as it seems. While in the long run the standard forms of regulation require very similar types of information, in the short run and certainly at the time the privatization contracts are being put together, they differ in terms of their effects on the optimal behavior of the company to be regulated.

QUALITY UNDER RATE OF RETURN REGULATION. Under rate of return regulation, overinvesting in quality may be a rewarding strategy for private investors. Indeed, under this form of regulation, prices are determined by the stream of revenue required to cover the allowed rate of return and the operational costs. To increase prices and, hence, cash flows—which, for all practical purposes, is what most companies aim for—the easiest solution is to increase the asset base. One way to do this is to overinvest in technological quality. In many countries in which airspace is still subject to limits to entry, buying a Concorde—to take an extreme case—rather than a Boeing or Airbus, and carrying the same number of passengers will increase the asset base without much of an impact on the demand for the service.

A limit to overinvestment clearly exists. This limit is determined by the interaction between price levels and quality on the one hand and willingness to pay on the other. For instance, a toll road concessionaire may have an incentive to overinvest in road quality, but only to the extent that the resulting toll level is consistent with the overall willingness to pay for road services and with the risks implied by the existence of competition in the form of alternative modes or routes. Moreover, the regulators always have the option of disallowing excessive investments from the rate basis and can get the service users to support this decision through consultation processes.[12]

---

12. Although if the regulatory framework allows this kind of discretionary power, regulatory conflicts will likely arise. Simple rules are often more efficient in a context in which the risks of political capture are present, as is the case in many developing countries.

QUALITY UNDER PRICE CAP REGULATION. Under a standard price cap regime, a subtle cut in quality can be a tempting way to cut costs. This is because under this regime, the difference between prices set for an extended period of time—three to five years—and costs is the main source of profit for the monopoly. Thus any reduction in quality that can lead to a reduction in cost implies a higher profit. While the regulator takes the bet that the service provider will try to cut costs by increasing efficiency, the most common way to cut costs turns out to be cutting maintenance. This, in turn, often means safety and environmental hazards in transport services, unless the regulator spells out clear quality standards regarding areas as diverse as reliability, interactions with users, safety, and environmental impacts.[13]

A quality-sensitive price cap has been suggested in the U.K. context. This can be done by allowing an adjustment in the productivity factor of the standard RPI–x formula for variations in quality. None of the regulators in the United Kingdom, for instance, has ended up relying on this approach. This is probably because determining levels of quality consistent with various aspects of efficiency is difficult for any regulator. Instead, regulators consider the international best practice indicators and use them as benchmarks when imposing quality standards.

COMPARING RATE OF RETURN AND PRICE CAP. While not much empirical comparative analysis of the impacts on quality of the two types of regulation has appeared yet, a few interesting points are worth noticing at this stage. The first is that often, particularly when it comes to technical standards, the actual benchmarks used as inputs in the definition of the price caps and in the calculation of the allowable rate of returns are the same. They just have a different interpretation. Under price cap, an international best practice benchmark is defined as a lower bound to be achieved by the service provider—because of a concern that quality will fall too low. Under rate of return, it can viewed as an upper bound for some indicators (as a way of avoiding overinvestment in service quality inconsistent with demand).

Second, under both types of regulatory regimes, regulators have been tempted, in too many cases, to include demanding input performance

13. A more thorough discussion would show that under a price cap system, rates of return on capital are also important, and in some cases, it may be a good idea for a firm to overinvest, just as in the case of the rate of return regulation. This arises because rates of return on capital are a key element of a price review.

criteria in privatization contracts. This is generally counterproductive. It is equivalent to creating a shadow management team for the private company run by the regulators. The temptation to micromanage is strong among new regulators, because they often were the managers of the same company when it was a public enterprise. Succumbing to that temptation reduces the interest of private operators to bid for the right to deliver a service. Even if some indicators of input quality are necessary when safety, health, and environmental considerations have to be taken into account, output performance indicators tend to be the better way to go about ensuring the desirable degree of quality without reducing the incentive for private operators to compete for the right to deliver the service. Particularly effective is when the regulator can use the information to promote yardstick competition.

PENALIZING FOR NONCOMPLIANCE. Monitoring for compliance is what most regulators will spend most of their time doing. But monitoring must be backed up by penalties if it is to have a disciplinary effect on operators. If monitoring is effective and the regulator sets the penalties correctly, the incentives are right for operators to supply quality levels specified by the regulator. In the absence of penalties, operators have no incentive to try and meet specifications, even if they are monitored.

The optimal penalty is equal to the operator's marginal net benefit from changing the quality offered (defined as the revenue minus the cost from the last unit of quality). But coming up with this precise calculation is often difficult. In the case of a rail franchise, various penalties can be conceived. One of these is to charge a penalty equal to the associated loss in consumer benefits from the service. This would require good information about demand. An alternative would be to set penalties equal to unit cost, taking the information required for this from train operators' accounts. The intuition is as follows: A train operator could obtain some gain by not running a scheduled train. If the regulator charges a penalty equal to those costs saved by not running the train, then the gain from not running the scheduled train is wiped out.

Pricing the Access to the Network

One of the most difficult technical areas for a regulator is designing network access rules. Yet these rules are an important component of policies for promoting effective competition in all segments of network industries, such as in railways and, more locally, in airports and ports, when some facilities have to be shared by an owner with competitors. They guarantee that competitors have access to the services of "potential bottleneck facilities" that are too costly to duplicate. Fair access rules to these facilities and prices will generally improve economic efficiency by easing competition in markets both upstream and downstream of the bottleneck. This is true whether the industries are vertically unbundled or separated or not.

In most cases, a vertically integrated industry competes in some markets with firms asking for access to the bottleneck facility it controls. Failure to design these access rules properly is one of the key reasons why the potential gains from restructuring network utilities are not achieved and/or shared fairly between the users and the owners of these essential facilities. When railways are concessioned, for instance, control of the rail network is generally left to private operators that may or may not also run trains on the tracks. This operator basically has full power to control entry onto the network. In cases in which the network operator also runs trains, it gets an unfair advantage, because it controls access to an essential facility for its competitors. This could result in low-cost operators being excluded from markets they would have otherwise won. The same story can be told for the concession of an airport to a company also owning an airline, or to a port operator also owning container terminals.

The three guiding principles for the regulator in this context are as follows:

- Allow a reasonable rate of return on the investment made on the network
- Ensure the effective coordination of all demands to access the network to guarantee efficient production and consumption of transport services
- Ensure that entry by new train operators is allowed and is fair.

Access charges are crucial to achieving these three objectives. Their assessment is controversial, however, because many of the costs of the bottleneck facility are common costs that need to be distributed across activities.[14] Even with a clear structural separation of the various activities of an industry, trade-offs exist between at least two types of efficiency: allocative efficiency (the best product mix for society) and productive

14. For a recent survey of the theory underlying the access pricing debate, see Valletti and Estache (1999).

efficiency (the cheapest cost for a given output mix). Because access prices are an integral part of the cost structure of firms downstream from the bottleneck facility, regulatory decisions allow firms to reduce the direct linkages between final prices and cost structures, influencing the two types of efficiency differently. Regulators may have to chose which type of efficiency to favor.

Things become even more complex when regulators have imperfect information on costs and on efforts to cut these costs by service providers. In this context, extra profits in many instances may have to be left to the network operator for situation-specific incentive reasons. The traditional way to analyze the situation is to figure out what the costs are based on, accounting for separation of the various activities. This is where controversy starts. First, the costs in question include both incremental costs (defined as costs directly related to the increase in production caused by the demand for access) and joint and common costs (costs incurred in the supply of a group of services that cannot be directly attributed to any one service and typically derive from economies of scope). The latter have to be allocated in the right proportions to the various activities, which is a very delicate operation.

The easiest way is to adopt fully distributed costs (FDC). Under FDC, the common and joint parts are allocated according to various measures: output shares (uniform markup per unit), directly attributable costs, and revenues or price-proportional markups. All these rules are mechanical, and therefore easy to implement, but are completely arbitrary. From an economic viewpoint, their main drawback is that they do not encourage cost minimization or account for demand. At the same time, they are relatively simple, familiar, and well understood, which may explain their popularity among practitioners. Moreover, under FDC there is a commitment to allow the network operator to recover its investments, which can be desirable in some circumstances.

An improvement in the accounting cost method is already an achievement, enabling reliance on more direct costs. It is still subject to criticism when based on accounting book values (historical cost accounting). In many occasions, the replacement cost of the bottleneck is different than its historic cost. Access charges based on historic costs can then send wrong signals to entrants, attracting too many inefficient firms, or discouraging potentially efficient ones. To overcome these difficulties, one can use current cost-accounting methods. They value and depreciate assets according to their current replacement cost. Typically, they involve the valuation of the firm's existing assets at the cost of replacing them with assets that serve the same function and are likely to incorporate the latest available technology. Such a forward-looking approach is fundamental for the calculation of long-run incremental costs (LRIC). They are often advocated as the best base for access charges if one wants to promote competition among entrants. Proponents of pure LRIC also believe that network economies of scale are not so pronounced, so that fixed cost recovery is not a problem. Once the incumbent faces revenue requirements, LRIC plus (often uniform) markups may be used.

The purely cost-oriented approach to access has raised many problems that are now leading to an alternative approach based on network usage. Among usage-based approaches, the most popular has been a formula called the efficient component pricing rule, also known as the Baumol-Willig rule. It is also the easiest to implement. It states that under some very specific assumptions (when final products are homogeneous and the market is contestable),[15] the access charge should be equal to the difference between the final price and the marginal cost on the competitive segment. This implies that the access charge should be equal to the direct cost of providing access plus the opportunity cost of providing access. The opportunity cost is the reduction in the incumbent's profit caused by the provision of access. While simple and attractive, the rule has serious conceptual problems that have led many to reject it, and it has generated a huge amount of research.

The most innovative proposal is to have a global price cap on the incumbent's entire range of products (Laffont and Tirole 1996, 1998). The rationale is that access services, in the eyes of the incumbent, are just a particular type of service. The bottleneck input should then be treated as a final good and included in the computation of the price cap. When a cap is properly set, the regulated firm is induced to choose the optimal Ramsey prices, without the need for the regulator to know the demand functions.[16] This is because the regulated firm can use its own private information about costs and preferences to set intermediate and final prices that satisfy the global cap, and these prices are the efficient one. The global cap is only an average value over a basket of

15. A contestable market includes free entry of firms and no unrecoverable sunk costs, so that firms are allowed to "hit-and-run."

16. The idea of Ramsey prices is that different prices should be charged to different users to allow the recovery of fixed costs, but this has to be done in a way that minimizes the distortions from marginal cost pricing. This is achieved by allowing the levying of a higher markup over marginal cost from the users who want the source the most.

services. The incumbent is left with the flexibility of increasing or decreasing all its prices as long as their weighted average satisfy the cap.

This proposal, while a major improvement over the efficient component pricing rule, still has its share of problems relating to the common ways of implementing price caps. For instance, the normal caps often used in practice only concentrate on final goods, which distorts price. The properties of global caps remain encouraging, even if they depend on strong assumptions. On top of the standard criticism that one can make about price cap mechanisms (weights in practice are based on realized outputs, which give a rate-of-return flavor), one additional concern is raised with global caps: the incumbent can engage in predatory practices. By raising the access price and lowering the final product price, the global cap can be satisfied while performing a price squeeze that damages new entrants.

In practice, however, the usage-based approaches are still at an early stage, and the cost-based methods are still dominating the regulatory decisions governing access prices. This trend will eventually change. Keeping up with progress is worthwhile in academic research, which should soon be able to provide guidance on reducing the distortions introduced by cost-based, regulatory decisions.

Public Service Obligations, Competition, and Equity

The desire and the ability to meet social objectives, particularly the needs of the poorest people, should not disappear with the introduction of competition and the widespread privatization of transport networks and services. Only explicit mechanisms are required to reconcile private participation in the sector with some of the social goals of governments. One such mechanism is to select the regulatory regime to ensure that the private operators have enough incentives to invest in activities they may otherwise have considered not profitable enough for the expected risk level, including the delivery of services to the poorest people. This is why the risk assignment issues discussed above are so important. Another mechanism is including the clear specification of the universal service obligations (USO) in the scope of monopolies' responsibilities.

Reconciling Social Concerns and Privatization: Universal Service Obligation

USO is an obligation imposed on the provider to ensure that anyone in its service area has access to an affordable, minimum level of a standard quality service bundle. A major source of concern for potential investors is that sometimes affordable means a price that may not necessarily cover the cost of delivering the service. Also, the precise definition of the range of services to be covered through the obligation varies. It may address spatial or geographical differences, specifying for instance that rural areas or inner cities must be serviced just like richer urban areas. One then says the USO is aiming to benefit high-cost customers. It may also focus on criteria more related to the income level of the potential users, or to specific demographic or institutional characteristics (such as retirees, schools, or hospitals). Low-income groups, for instance, cannot always afford the connection costs to a water main at prices that other income groups can. Moreover, they typically cannot borrow as well—because of capital market imperfections in many developing countries—which further limits their access to these services.

As is the case for any regulation that puts pressure on costs, the private operators may cut quality to recover the additional financial burden resulting from the need to invest to meet USO. For instance, train and bus operators can always marginally decrease safety through an excessive use of trains or buses to meet harsher USO. Private infrastructure suppliers could also be tempted to decide the timing of the investment requirements on their own. They could accelerate investments for access in rich suburbs and slow down expansion in poor inner cities, while meeting the overall coverage increases demanded by the government in its privatization contracts.

International experience suggests that for many developing countries, the most effective way of clarifying USO may be to translate it into specific and transparent targets consistent with the government's overall social and economic objectives for the sector. Every privatization contract should have a detailed appendix that covers this issue. This appendix could include some specific guidelines on the speed and location of investments, on the speed of cost recovery, and on the transparency of required subsidies. One should assess the financial consequences of these guidelines against the financial and pricing conditions proposed by the concessionaires in their bids as a way of making these goals a reality.

The challenge resulting from the idea that access to an infrastructure service met through USO is a paying, yet affordable right implies that USO coverage, quality, pricing, and financing mechanisms have to be decided jointly. The specific targets suggested in the contractual arrangements are crucial in determining the financial viability of the service. This is particularly important for developing countries. When the concessionaires have an obligation to meet any reasonable demand, this obligation has to be translated into an investment plan with targets spread over time—for as long as 10 years in some contracts—to allow a distribution of risks during that period.

Pricing and Financing of Public Service Obligations

When the government demands that an operator deliver services at below-cost prices, finding out if the government will provide subsidies to pay for the difference seems appropriate. Or will the firm be asked to find its own ways of financing the gap? When the government decides to subsidize the costs of delivering public service obligations, it must also pick a subsidy design and develop a monitoring system that will allow the regulator to ensure that the operator is not trying to deliver more than needed to generate excess profit from subsidy payments. This is a risk, for instance, in public bus services or in subways in which the government usually lets the volume of passengers drive the subsidy levels.

When governments face tough fiscal constraints and are unable to finance subsidies, they sometimes impose uniform pricing requirements as a complement to USO, explicitly impeding price discrimination and implicitly imposing cross-subsidies. The resulting average pricing approach may be sufficient to ensure that the operator does not make losses in the delivery of the service. However, this implies that high-cost users (those who need new connections, often the poor) are subsidized by low-cost users (those who already have connections, often the middle and upper class or businesses). This means pricing one activity above-cost to finance another activity priced below-cost. The main problem is that traditional cross-subsidies have proven to be opaque and inefficient in the allocation of investment decisions, thus opening opportunities for corruption or unfairness.

Unfortunately, international experience suggests that the cross-subsidies are often poorly designed, and they typically fail to meet their financing and expansion goals. Economic advisors typically suggest rejecting the use of cross-subsidies and paying a targeted subsidy financing as a lifeline to the poorest beneficiaries of the USO, but this only works if the government has the ability and resources to manage a lifeline program, which is often not the case. An increasingly common alternative in the power and telecommunications sectors is to consider financing USO through sector-specific levies on either users or operators. This avoids the harsh demands or distortions on the poorest people due to cross-subsidies. Some countries are also now discussing the possibility of introducing various types of sector-specific funds. These are all new ideas that must be monitored before making any serious assessment.

To minimize financing requirements, reformers are also trying to minimize the costs of USO from the beginning of the reform process. This can be done in several ways. One approach that Peru is currently considering is an auction of the rights to use a spectrum in the telecommunications sector. The service and related obligations will be awarded to the bidder asking for the lowest subsidy to meet the USO. A similar idea was initially considered for the concessioning of freight rail service in one of the poorest regions of Brazil. A reallocation of idle public assets from regions that could be concessioned for a positive profit to that poor region eliminated the need for subsidy in that case. This asset reallocation was sufficient to result in an auction for the highest price, rather than at a minimum subsidy. Sometimes focusing on the choice of technology to meet the obligations can provide affordable solutions. Also, working out alternative institutional arrangements can ease the financing constraint and improve access to networks. In Argentina, electricity is being supplied to local cooperatives responsible for paying a collective bill, which reduces the risks of nonpayment for the distribution company. However, technological alternatives are not always possible. One needs to identify alternative financing strategies to ensure the financing of the USO without threatening the financial viability of the activity.

Regulatory Institutions

An essential element of an effective regulatory framework for privatized transport infrastructures and services is to place the responsibility for regulation in an agency with the required independence, autonomy, expertise, and accountability. This agency must protect the interest of both users and investors. It must do so in a fair and transparent way. This, in turn, requires skills, independence, autonomy, and accountability. In practice, creating an institution with all these skills seems to be a particularly difficult challenge, and most countries are having a tough time adopting the right principles in the creation of their institutions. The standard recommendations are, however, straightforward. This section summarizes the major decisions a government has to make when deciding to consider creating a regulatory agency.[17]

The Regulator's Sectoral Breadth of Authority

The first concern to address is the sectoral breadth of authority. Governments can organize entities in one of three main ways as follows:

- *Industry-specific:* separate agencies for roads, rail, electricity, and so on.
- *Sector-specific:* separate agencies for a group of related industries, such as an energy regulator for electricity and gas or a transport regulator for rail, roads, and ports. Argentina and Peru, for instance, have a transport regulator and Brazil and Mexico soon will have one, too.
- *Multisectoral:* a single regulatory agency for all or most infrastructure sectors as in the case of state-level regulators in Australia, Canada, and the United States and national regulators in Costa Rica and Jamaica.

Most experts seem to argue that the broad base of a multisectoral agency offers advantages over the alternatives. Not the least is the opportunity to share regulatory resources (regulatory economists, lawyers, and so forth) for countries with limited regulatory capacity. However, it also allows for greater resistance to efforts by specific sectors to control the regulatory decisions (industry capture) or to political interference with these decisions. It also has the advantage of allowing consistency of decisions across sectors because the various interest groups have more opportunities to control each other.

Many experts also recognize that in the case of the transport sector, a sector-specific agency may make sense. In most countries, a transport ministry or a ministry of public works controls the transport sector, and the transition to a specialized regulatory agency is often easier if the regulator has a similar sectoral coverage.

In practice, thinking of the optimal design of the institution may make sense when all reforms have been put in place, but recognizing that not all sectors are equally easy to reform is important. This means that some sectors will require independent regulation sooner than others. For instance, transferring the Brazilian railways sector to private operators took about a year, and it is taking much longer to transform airports or ports. Does this mean

---

17. For a recent expert overview, see W. Smith (1996).

Brazil should wait to create an independent regulator until all sectors are privatized to the extent announced in the reform program? No, of course it does not. Brazil is already building up its regulatory capacity and thinking how it will fit into the future transport regulatory agency. What matters is having a vision of what the ideal agency will be and working out a politically sustainable plan for its implementation, if needed, sector by sector.

The Desirable Qualities of a Regulator

While the important issue of a transport regulatory agency's desirable qualities is complex, and this brief overview cannot do it justice, a few minimum requirements have to be addressed for regulation to be successful. The following are the main desirable characteristics usually recognized for a regulator:

- Independence with a reasonable amount of discretionary powers
- Autonomy and expertise
- Accountability.

Independence means that ideally, regulators should be at arm's length from political pressures, in particular from ministries. In other words, leaving regulators within the ministry under direct control of the minister—who, after all, is a political appointee—is generally a bad idea. But regulators should stay at arm's length from the regulated enterprises as well, whether these are private or public, because it is not in the consumers' interest to have the utilities influence the regulators' decision in some arbitrary way either. The other minimum requirements of independence are ensuring that the regulators are appointed on the basis of professional rather than political criteria and ensuring that the appointees are protected from arbitrary removal during their appointment term. Ideally, the process should involve both the executive and the legislature.

To achieve independence, the regulator must have more than just an advisory role (advisors by definition make no decisions) and must enjoy a certain degree of discretion in its decision by rule. The regulator should indeed have enough power to make the right decisions without risks of interference. The regulator having too much discretionary power is, however, a risk. The consequences of this excess of discretion are often easy to assess. An independent regulator with strong discretionary powers that wants to favor investors will have to face user complaints. This will result in difficult political situations. The regulator that uses large discretionary powers to favor users will increase the perception of regulatory risks, which in turn will increase the cost of capital in the sector. This in turn will eventually hurt users as well, because these cost increases are passed on through to the tariffs. One way to minimize these risks is to spell out clear rules in the contract and in the chart creating the regulatory agency.

Autonomy covers various issues. Agencies need to have access to their own funding sources. Relying on budgetary transfers decided by politicians is often viewed as a threat to the regulators' independence, because an easy way to reduce the effectiveness of a regulator would be to cut its budgetary allocation. Levies on the regulated firms or the consumers of the regulated services are the most common alternatives and can be viewed as user fees to be paid for the protection services provided by the regulators. Ideally, the government should determine the levies annually, based on budget proposals the agencies submit.

However, autonomy has to be more than just financial. It should also mean that the regulator can recruit the best and that the size of an agency staff should not be determined by the number of people who have lost their job through the privatization process. It should depend on the specific tasks assigned to the regulator under the particular regulatory regime, and a large staff is not necessary to achieve this. Having too much staff can lead to interference with the commercial operation of the regulated firms. Achieving staffing autonomy will often require an exemption from civil service salary and recruitment rules. It may also imply that these agencies have to be able to recruit external consultants when the required skills are not available locally to address specific needs. Access to local expertise in the relevant economic, accounting, and legal principles is indeed often a challenge; thus, regulatory capacity will be limited if subcontracting of some activities is not allowed.

Finally, autonomy in the monitoring of compliance and enforcement deserves to be highlighted because it requires that specific instruments be assigned to the regulators. To be effective in this role, the regulator must be able to impose penalties according to clearly defined rules. Ideally, penalties should be linked to the damage imposed on users or competitors and hence returned through rebates to the victims.

Accountability requires transparency in the decisionmaking process, which unfortunately too often is counterintuitive to many bureaucrats. It also requires an operating environment subject to simple and clear procedural rules, including stipulated deadlines for reaching decisions, detailed justifications of decisions, nonpolitical reviews of decisions, opportunities for all concerned parties to be heard through public hearings (and hence greater interactions with consumer defense groups) and venues for appeal, and provisions permitting the removal of regulators in the event of proven misbehavior. One of the key elements influencing the level and type of accountability is the decision on the number of regulators. In most cases, a regulatory commission with three to five members should be preferred over a single regulator. It increases accountability to the extent that each member of the commission ends up monitoring the others. This is now widely recognized even in the United Kingdom, which is the original model for the single regulator approach.

The Division of Labor between the Regulator and the Government

The effectiveness of the regulator depends to a large extent on the clarity with which the sector responsibilities have been divided between the regulator and the transport ministry or secretary as well as with any other government institution with potentially overlapping responsibilities. Table 2.3 provides practical guidelines. These are, of course, suggestions that are subject to adjustment to the local circumstances, and they entail a certain

Table 2.3. *The Division of Labor Between Regulator and Government*

Features	*Government*	*Regulator*	*Other entities*
Legal framework and sectoral policy	•		
Planning	•		
Privatization design	•		
Taxes and subsidies decisions	•		
Concessioning and procurement auctions	•		
Pricing		•	
Control and penalties		•	
Technical regulation		•	
Quality standards		•	•
Environmental regulation		*	•
Safety		*	•
Health		*	•
Antitrust policy		**	•

* Requires coordination to assess impact on costs.
** Requires coordination and functioning of industry to assess financing requirements.
Source: Authors.

degree of subjectivity. One could argue, for instance, that the regulatory agency should be involved in the design of privatization and concession contracts because it will hold the main responsibility for implementing them. However, agencies are seldom in place on time for this to happen.

Conclusions

Far from eliminating the necessity to regulate, the experience of the past 10 to 15 years with the increased role of the private sector in transport has demonstrated the importance of effective regulation and effective regulatory institutions. The need for economic regulation in the sector has tended to be enhanced by the existence of natural monopolies, the limitations of competition for the market, the difficulties of managing the information asymmetries between the private operators and the regulator, and the increased role of project finance in the expansion of the transport network, as well as the tricky interactions between risks and regulatory choices and decisions.

Because most of the infrastructures are operated as local monopolies, regulators have to provide themselves with the tools required to ensure that the increased private sector participation benefits the majority of the society and not just the investors, while protecting the commitments made by the governments to these investors and operators. Economic regulation must balance incentives and risks in a way that stimulates efficient and fair behaviors and maintains the cost of capital to levels consistent with reasonable participation of the private sector in the financing of transport infrastructure and services.

Recent developments in the theory of regulation discussed in this chapter address many of the informational problems that plagued the previous generation of regulatory regimes. They have led to a reliance on more commercial contracts between operators and the government to specify many key regulatory requirements, replacing the somewhat rigid and sometimes excessively ad hoc regulatory rules that emanated from regulators that too often lacked concern for cost-effectiveness, efficiency, and customer orientation. This offers opportunities to improve relations between users and operators.

Sadly, the opportunities these new theories offer, as well as these instruments, are only as good as the regulators responsible for implementing them. The rest of this book shows the choices, the trade-offs, and the challenges of transport privatization. It provides the information that is necessary for success in each sector, but it does not and cannot have the ambition to provide a sufficient condition for success. Success will be achieved through the political commitment of those asked to implement and support the changes. Few books will be able to provide guidance on this, but here is a start for the transport sector.

3

Airports

Ofelia Betancor and Roberto Rendeiro

Airports around the world are vital infrastructures that guarantee connections. Traditionally they have been essential transport interchangers between air and surface modes. Central or local governments, and in some instances, a branch of the army, have owned and operated them. Supporting this type of ownership was the common belief that airport infrastructure was a public utility. Because of public budget constraints and efficiency concerns, however, this model has been reconsidered. Airports are no longer regarded as simple interchangers—they are businesses and, as such, have to be managed. The traditional model seems unsustainable, and involving the private sector in airport activities appears to be the best alternative. Today, the range of possibilities for private sector involvement in airports may be as wide as the range of airport activities themselves.

Airport Services

Airports are complex, multiproduct enterprises. Each airport comprises one or several runways, a set of aprons and taxiways, a terminal building that separately processes passengers and freight, and a control tower. Each of these parts develops specific activities that, once combined, allow the interchange between air and land transport modes. This combination is a complex system that serves a wide range of needs related to the movement of people and things worldwide. Its development depends on four crucial elements: passengers and goods that circulate through its terminals; its physical, social, and economic environment; its nature as a productive, business-generating unit; and the agents that operate within it, mainly airlines and commercial service franchises.

The Multiproduct Nature of the Activity

The activities carried out at airports may be classified into three distinct groups: essential operational services and facilities, handling services, and commercial activities (see table 3.1) (Doganis 1992). The first two are commonly referred to as aeronautical services, while the latter are considered nonaeronautical.

Essential operational services include the air traffic control system, meteorological services, telecommunications, police and security, fire, ambulance and first aid services, runways, aprons, taxiways, and grounds and building maintenance. These activities determine the safety of airport operations and, hence, are considered essential to the airport business. Handling services refer to a great variety of activities. We can distinguish between those that are directly related to the aircraft (ground and ramp handling), such as cleaning, providing power and fuel, and

Table 3.1. Classification of Airport Activities

Aeronautical or airside services		*Nonaeronautical or landside services*
Operational	*Handling*	*Commercial*
1. Air traffic control	1. Aircraft cleaning	1. Duty-free shops
2. Meteorological services	2. Provision of power and fuel	2. Other retail shopping
3. Telecommunication	3. Luggage and freight loading and unloading	3. Restaurants and bars
4. Police and security	4. Processing of passengers, baggage, and freight	4. Leisure services
5. Fire, ambulance, and first aid services		5. Hotel accommodations
6. Runway, apron, and taxiway maintenance		6. Banks
		7. Car rental and parking
		8. Conference and communication facilities

Source: Adapted from Kapur (1995).

loading and unloading luggage and freight; and those that are traffic related (traffic handling), such as processing passengers, baggage, and freight through the terminal building. Finally, commercial services involve a large variety of different activities that may either be located at the terminal building or around the airport. Examples of the myriad of activities included in the nonaeronautical set of airport operations are duty-free shops and other retail shopping, restaurants and bars, leisure services, hotel accommodations, banks, car rental and parking services, and conference and communication facilities.

Airlines are also involved in the commercial side of airport activities. Carriers usually need an office at the airport, which the regulator must consider. Under scarce space conditions in terminals, the relevant question is how to ensure a place for every carrier. A transparent, competitive process should ensure that major airlines receive space; in some cases, these airlines may also represent more minor airlines. The same transparent process should be in place for the assignment of space for VIP lounges.

The classifications in table 3.1 do not apply to all airport activities. Sometimes the criteria that allow one type of service to be separated from another become blurred. Aeronautical or airside activities focus on the operation of aircraft and the movement of passengers and freight, while nonaeronautical or landside activities are connected to commercial operations occurring in the terminal and on airport land, usually under a concession contract. Any concession that relates to aircraft or traffic handling would share some features with both aeronautical and nonaeronautical services. Fuel concessions and passenger and freight handling, when provided by an airport agent, are examples of activities that would not fit into table 3.1. Therefore, the classifications shown in the table should be regarded as tentative.

Airport Revenues

With the assumption that the sorting of airport activities is no longer a problem, revenues arising from these services are also classified as aeronautical and nonaeronautical. A relationship exists between airport size and revenue generation sources: bigger airports are more capable of exploiting commercial activities and hence obtain more revenue from this source. In contrast, small airports tend to be almost entirely dependent on aeronautical revenues. Empirical evidence for this type of relationship in regard to Spanish airports is shown in table 3.2 and corresponding figure 3.1.

Table 3.2. Airport Size and Revenue Sources: The Spanish Case, 1997

Airports	Passengers (thousands)	Aeronautical revenue/total revenue (%)	Nonaeronautical revenue*/total revenue (%)
Largest airports			
Madrid/Barajas	23,122	58	42
Palma de Mallorca	16,449	64	36
Barcelona	14,561	60	40
Average for largest airports	18,044	61	39
Large airports			
Gran Canaria	7,927	68	32
Tenerife Sur	7,438	71	29
Málaga	7,190	55	45
Alicante	4,398	56	44
Lanzarote	4,005	77	23
Ibiza	3,528	61	39
Fuerteventura	2,440	71	29
Menorca	2,232	62	38
Tenerife Norte	2,042	63	37
Bilbao	1,970	65	35
Valencia	1,912	60	40
Sevilla	1,543	57	43
Santiago	1,283	61	39
Average for large airports	3,685	64	36
Medium airports			
Almería	714	68	32
La Palma	696	67	33
Asturias	595	54	46
Vigo	556	67	33
Reus	518	75	25
Gerona	507	66	34
Jerez	453	56	44
Granada	447	59	41
La Coruña	398	62	38
Melilla	352	80	20
Average for medium airports	524	65	35
Small airports			
Pamplona	288	71	29
Zaragoza	244	71	29
Santander	204	72	28
Valladolid	191	82	18
San Sebastián	173	70	30

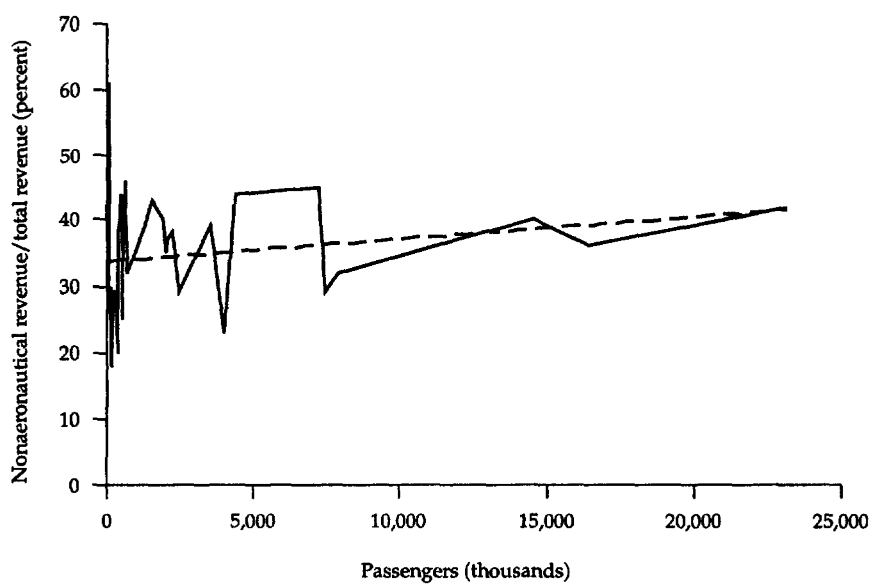
(table continues on following page)

Table 3.2 continued

Airports	*Passengers (thousands)*	*Aeronautical revenue/total revenue (%)*	*Nonaeronautical revenue*/total revenue (%)*
Vitoria	145	77	23
San Javier	108	71	29
El Hierro	97	39	61
Salamanca	44	74	26
Badajoz	18	64	36
Average for small airports	151	69	31
Total Average		65	35

* Including handling.
Source: Aeropuertos Españoles Navegación Aérea data.

Figure 3.1. *Airport Size and Revenue Sources: the Spanish Case, 1997*



Source: Authors.

According to Doganis (1992), when an airport reaches the 10 million passengers threshold, commercial revenues represent between 50 and 60 percent of total income. U.S. airports are an exception, however, with 70 to 80 percent of total income typically coming from commercial revenues.

Such differences are primarily due to U.S. airports leasing out terminals, hangars, and other facilities to airlines.

More relevant is the relationship found between the type of ownership and revenue generation. The arrival of the private sector into airport operations has led to what is called the commercial airport model, which goes beyond a traditional airport to regard infrastructure as a business opportunity. Meetings, visitors, employees, local residents, and local businesses and industries are also important potential customers for commercial services at the airport. From this point of view, the greater the involvement of the private sector in airport activities, the greater the importance of nonaeronautical sources of revenue. As table 3.3 shows, this is what is occurring, except at regional airports, which for this sample are mainly located in the United States.

Airport Demand

Demand for basic airport services such as aircraft landing is directly influenced by the air transport market, which in turn depends on trip

Table 3.3. *Traffic and Revenue Distribution: Selected Airports*

Average	*Government department*	*Public corporation*	*Regional*	*Public-private*	*Private*
Annual aircraft movement (thousands)	78	165	391	169	188
Number of passengers (millions)	6.6	11.9	28.4	12.0	11.1
Airside revenue as percentage of total revenue	70	50	36	62	43
Landside revenue as percentage of total revenue	30	50	64	38	57

Note: Different airport ownership structures are defined and analyzed later. Selected airports include the following: *Government department*: Buenos Aires, Santiago, Mexico City, Quito, Libreville, Nairobi, Budapest, Athens, Gothenburg, New Delhi, Hong Kong, Bangkok, and Kuala Lumpur. *Public corporation*: Sydney, Auckland, Singapore, Rio de Janeiro, Amsterdam, Madrid, Vancouver, and Montego Bay. *Regional government*: Washington, Boston, Chicago, Pittsburgh, Atlanta, Dallas, Miami, Orlando, Paris, and Basel-Mulhouse. *Public-private*: Toronto, Vienna, Rome, Copenhagen, Zurich, and Yaounde. *Private*: Heathrow, Gatwick, Stansted, Aberdeen, Edinburgh, Glasgow, and Southampton.

Source: Kapur (1995).

purpose. Hence, it is considered a derived demand. Demand for landing is generally price-inelastic (see, for instance, Morrison 1982), because airports usually do not have a local competitor and airport charges represent a small proportion of the direct operating costs of airlines. Doganis (1991) reports approximately 5 percent for airport and en route fees for International Civil Aviation Organization (ICAO) airlines.[1]

As Walters (1978) noted, air transport demand is subject to two motivations—business and leisure. Therefore, we can distinguish between at least two distinct types of airline consumers: business and leisure travelers. Each group may also be divided into different subcategories. For instance, some business passengers need complete flexibility and others travel according to plans. For leisure customers, some are people traveling to vacation resorts and others are traveling to visit relatives or friends.

These groups behave differently in the market. Leisure travelers are quite price responsive while business passengers tend to be less so, although they are not totally price inelastic. Business travelers are also influenced more by the convenience of schedules, because they usually book their tickets at the last minute and frequently alter their bookings. Business trips are concentrated in the early morning and late evening hours, while leisure traffic principally appears on weekends and holiday seasons.

Consequently, airport service demand is characterized by peak and off-peak fluctuations, which can be found by day, by week, and by season. Airport capacity is strained by this peak nature of demand. Furthermore, when taking the spectacular growth in the air transport sector into account, analyzing airport capacity becomes essential.

Capacity Constraints

The term "capacity" refers to the ability of an airfield component to accommodate aircraft movements. It is expressed in terms of operations per unit of time (usually per hour). For instance, the hourly capacity of the runway system is the maximum number of aircraft that can be processed in an hour according to a set of specified operating conditions.[2] Therefore, when evaluating airport capacity, one should study terminal building and runway system capacities individually, although the latter is usually considered the main

1. The ICAO is an intergovernmental institution that was created at the Chicago Convention in 1944.

2. This concept of runway capacity, called saturation capacity, is presented later.

determinant of total system capacity (Ashford and Wright 1992). Four main factors affect runway capacity: air traffic control, demand, meteorological conditions at the airport, and the design and configuration of the runways.

Two basic concepts of runway capacity may be applied: practical capacity and saturation capacity. Practical capacity relates to the number of operations that can be done in a period of time without imposing an average delay that exceeds a reasonable, preestablished level. For instance, the delay for departing flights averages four minutes during peak hours. Saturation capacity refers to the maximum number of aircraft that can be served in a given period of time under continuous demand conditions.

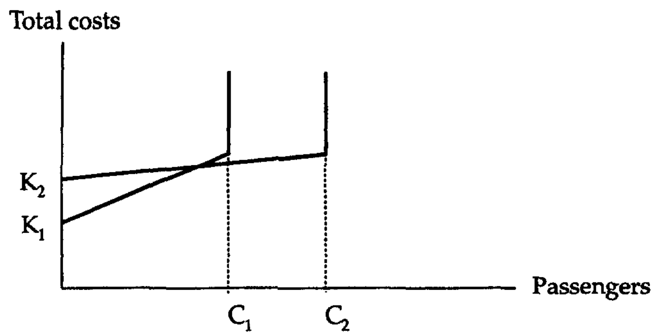
Airports are productive units whose capacities can only be increased through the incorporation of large, indivisible units. If runway capacity at a given airport is equivalent to a maximum number of *n* airplanes per period, and the airport operates below that level, the cost of operating an additional plane would be equal or close to zero. If the same airport operates at full capacity, however, increasing traffic would require the construction of a new runway. Therefore, if traffic volume at peak periods increases sharply, obliging the construction of another runway, that runway would be underutilized during off-peak hours. Fluctuations in demand for airport services and investment indivisibilities inevitably lead to excess capacity, which has significant repercussions on the cost structure of the airport industry. Peak period pricing, however, may help to lessen that problem and allow for more efficient allocation of capacity.

The shape of the cost curve for runways exhibits a positive slope for traffic volumes below available capacity. Once capacity is exceeded, the cost grows asymptotically. This is known as capacity cost and its behavior is shown in figure 3.2.

Terminal building capacity is becoming more important as the emerging role of the private sector has given nonaeronautical airport services a greater weight. Commercial and other activities carried out in the terminal building require large spaces. This capacity can be evaluated by considering two important variables: level of service and volume of service. Level of service is closely linked to quality. Space, waiting time, users' comfort, and treatment by airport staff are all determinants of quality. Evaluating these factors is subjective. For this reason, most studies use time of service and level of congestion as proxies for this variable. Volume of service, the second parameter under consideration, refers to the number of users that can be served given a selected level of service.

From the airport's point of view, establishing an adequate level of service is important, because the time that passengers waste while waiting in

Figure 3.2. *Runway Costs Functions*



K_1 Small runway capital cost.
K_2 Large runway capital cost.
C_1 Small runway capacity.
C_2 Large runway capacity.
Source: Walters (1978).

line renders a large number of resources useless. For instance, the greater the time required for check-in procedures, the less time is available to engage in last-minute shopping in the commercial area of the airport.

A shortage of capacity at airports translates into increasing congestion and delays. The immediate consequences for users are increasing costs and decreasing quality of service and safety. Providing additional capacity to meet demand requirements, however, has important implications for the airport cost structure.

An alternative mechanism for meeting demand is allocating flight and gate slots. Flight slots refer to landing and departure times and gate slots concern terminal utilization. When allocated, both types have to be jointly considered so that landing planes aren't delayed by waiting for a gate to become available. Traditionally, incumbent airlines have been the *de facto* proprietors of slots. These airlines have been using the gates for so long that almost all national laws recognize their property rights, or grandfather rights, based on regular utilization. This is the criterion that the International Air Transport Association (IATA) members recommend and accept.

Supporting grandfather rights in a deregulated air transport environment aimed to increase competition makes little sense. They are an efficient barrier to market entry. In that case, other methods may allocate airport slots. A second possibility is a slot auction in which airlines bid for a slot or a combination of slots. This mechanism ensures that the airport authority gets the highest possible price. Auction implementation is complicated, however, when grandfather rights are in place. Airlines that have these rights would not submit to an auction unless they were obliged by law to do so. Typically, in this situation, only newly created slots or those lost under a use-it-or-lose-it principle would be available for auction.[3] Furthermore, the allocation process requires that access to other airports also be considered. Slots are so vital for airlines that they have triggered important international alliances such as the one intended between British Airways and American Airlines that recently became subject to the scrutiny of the European Commission. The commission demanded the disposal of 267 slots at Heathrow and Gatwick airports as the price for approval of the alliance. The prospective partners, in turn, have requested that they be allowed to sell the slots.

Airport Costs

Airport costs fall into two categories: those related to the terminal building and those associated with the runway system. The first category depends on passenger flows in the terminal building and the second is determined by the number of aircraft processed. Empirical evidence points to the existence of economies of scale in landing operations, which means that as an airport increases its traffic, the cost per unit of traffic declines. By contrast, there are decreasing returns to scale when handling passengers inside the terminal. The required time to process a passenger through a terminal increases with airport size. Hence, the optimal dimension of an airport depends on a delicate equilibrium between both of these elements (Walters 1978).

Labor, capital, and other operational costs compound airport costs. Among Western European airports, staff or labor cost is the largest item, representing an average of about 42 percent of total costs (Doganis 1992). In a few cases in which airport authorities are involved in activities usually undertaken by concessionaires, such as handling services, this percentage may rise to 65 percent. The second major heading is capital charges (interest and depreciation). For most European airports, this figure ranges from 20 to 35 percent. In contrast, the cost structure of U.S. airports appears to be quite different. Staff costs may reach an average of 22 percent, and capital charges increase to 44 percent of total costs. These differences can be explained by

---

3. This is the method the European Union has selected, and perhaps one of the main reasons that may explain why competition has not flourished across Europe.

the different ways that both groups operate. For instance, the contrast in labor costs can be explained by the common practice at many U.S. airports of renting terminals and other facilities to airlines, which sometimes may even own the facility. The fact that concessionaires usually carry out handling activities also contributes to this explanation. Regarding differences in capital charges, it should be noted that U.S. airports have frequently used capital bond markets to finance development, while Europeans have been more dependent on government budget allocations.

Externalities in the Airport System

When users of airport infrastructure impose a cost-benefit upon non-users (or even upon other users on the system), an externality exists, because airport users are not bearing all the costs generated by the services they require. Most studies have been devoted to noise (see Walters 1978; Nelson 1980; Levesque 1994). Pollution and congestion, however, are other negative externalities that cannot be neglected from the regulator's point of view.

The main economic problem with externalities is quantification and subsequent valuation. For example, a highly sophisticated technology is required to measure aircraft pollutant emissions. For noise measurement, acoustic specialists have devised ordinal scales constructed as weighted averages of the high-frequency peak noise and the number of times aircraft noise is heard. Examples are the Noise Number Index in the United Kingdom, the Composite Noise Rating and the Noise Exposure Forecast in the United States, and the Isosophique in France. Congestion, a type of negative externality imposed upon other users in the system, must be linked to airport capacity to be measured. Capacity is a given, and the international standard for aircraft movement is that the average delay does not exceed four minutes. Longer aircraft waiting times indicate a congestion problem, which can also be measured with weighted averages.

Once the measurement task is completed, the emerging question is that of valuation. Because we are considering costs imposed by users upon non-users, we have to take into account people's judgments about damage suffered. Although the subjective nature of such a judgment makes valuation extremely difficult, economists have developed several tools that allow for a valuation that is more or less accurate.[4]

---

4. A good review for those interested in externalities is Christensen and others 1998.

Walters (1975) has argued that almost all of the problems that airport noise cause are reflected in lower property values. Hence, *ceteris paribus* comparisons between noisy and quiet houses should provide a market valuation of quiet. The price of noisy houses near an airport may be 30 percent less than the equivalent in quiet areas. This has been the subject of considerable controversy. It is perhaps the most dominant negative externality, and some airports have attempted to solve or lessen it by restricting night operations or establishing landing charges according to the noise level. This last option internalizes the negative externality: by paying for the disturbance incurred, airlines bear the true social costs. Quantification and valuation problems remain, however.

Coordination of Activities: Air Traffic Control

Privatization schemes have usually left out air traffic control (ATC), which has been under government control. Nevertheless, this trend is changing. For instance, New Zealand has corporatized the ATC; a limited liability company with two shareholders, the Ministry of Finance and the Ministry of State-Owned Enterprises (see box 3.1), operates it. The Canadian government went even further in 1996, selling its ATC to a private operator, Nav Canada, which is subject to an economic regulatory regime (see box 3.2). Most ATC systems, however, have not been privatized because of the fear that commercial pressures could compromise safety standards. Opponents of airline deregulation also expressed this fear. This situation includes two possible views (Chalk 1993): the market failure view and the market response view. According to the former, privatized airlines or ATC private operators face negative financial and safety incentives, suggesting that they could be inclined to reduce their safety expenses to increase profits. The second view suggests that because reduced safety can be observed in the form of accidents, consumers will use this as an indicator of an operator's level of safety, therefore penalizing negligent firms, possibly forcing them to leave the industry. For the airline industry, enough evidence supports both views (see, for instance, Rose 1990; Borenstein and Zimmerman 1988). Both the market failure and market response views influence the actual industry safety levels, indicating that safety regulation is necessary, although in practice it has been imperfect and complemented by market mechanisms. When privatizing ATC systems, one should take that experience into account.

Box 3.1. *ATC: The Case of New Zealand*

During the 1980s, the public provision of services in New Zealand underwent radical reform. The air traffic control service, which was operated by the Civil Aviation Department, was transformed into a commercially oriented corporation. The new organization, Airways Corporation, had to assume the responsibility for its management. Two of the proposed objectives were coverage of costs and provision of services required by users.

The Airways Corporation needed to be financially autonomous, which compelled it to adopt a new performance philosophy. Before the introduction of the new commercial orientation, service managers tried to please the politicians who controlled funding. The new approach, however, focused more attention on users, who were frequently consulted in regard to fare structures, the introduction of new technologies, and safety measures. The new approach permits decision making to be more flexible regarding the services users need.

Opponents of the ATC corporatization program were concerned about safety, detecting a conflict between safety and commercial goals and assuming that standards would decrease as a result of profitability pressure to reduce costs. Nevertheless, in New Zealand, it appears that the market may discipline such behavior. Conforming to certain standards is necessary so that consumers do not switch to other transport modes to avoid airports or air carriers with unsafe reputations.

The main achievements of the ATC corporatization in New Zealand include the provision of services at a substantially lower cost, the reduction of fares, the service improvements allowing users to obtain cost savings, and the adoption of new technologies and services. An important explanation for the success of this approach is that the board of directors includes people who are experienced in both the public and private sectors. In addition, the politicians and the government were resolute in their commitment to change. Also key to the transition was the government's recognition that a commercial approach can provide more efficient service. Indeed, the private sector in New Zealand accepts Airways Corporation as one of the country's best managed public enterprises.

Privatization and Regulatory Trends

Tradition has regarded airports and airlines as integrated and important parts of the national air transport system. Both were considered public utilities. In welfare terms, the benefits to society stemming from operating these services would always compensate for eventual financial losses, and would thus justify corresponding subsidies.

This model considers operational and handling activities essential to the airport business, while commercial activities play a secondary role. Airports are aimed to facilitate the interchange of transport modes, not to exploit passengers' willingness to pay for things they might buy at other, more adequate places. Conveniently, airport property, assets, and management are always in public hands, with only commercial activities some-

Box 3.2. ATC: The Case of Canada

In the Canadian case, establishing a commercial approach to air traffic control was the result of user demand for more efficient service. At the same time, corporatization of the ATC was also part of a government initiative to promote the modernization of transport infrastructure and a more rational use of resources in Canada.

Problems associated with the service included users not paying the true value and managers who were subject to rigid public rules that lacked the flexibility required by market conditions. In addition, the labor force was out of proportion with service needs. Finally, the slow, bureaucratic process for approval of investments made it difficult to incorporate new technologies into the system in accordance with market needs.

For all these reasons, in 1995, the Canadian government announced the commercialization of its air navigation system, establishing a set of principles to be assumed by the new operator. Among these were preserving and promoting aviation safety, improving the efficiency of the system, allowing access to all users, providing service to remote regions, complying with international obligations, and operating the service under a commercial approach with the goal of recouping all costs.

In turn, the government committed itself to developing regulations that would not affect the company's commercial interest. Nevertheless, it adopted some regulatory measures to prevent the firm from exploiting its monopoly power. It aimed to promote efficiency through the application of self-regulatory mechanisms that would give consumers enough protection at the lowest regulatory cost. The government also established a consultation procedure to maintain equilibrium among participants and minimize disputes requiring third-party intervention. Finally, it ensured noninterference between social and financial objectives. Such a regulatory structure aimed to protect users' interests while guaranteeing enough flexibility for the firm to maneuver in a commercial environment.

A report by Corporate Services of Canada regarding the commercialization of the air navigation system noted that the experience was successful for all parties involved. The industry maintains its safety level and the system responds to demand and technology changes more efficiently. Travelers and users benefit from more efficient service. The government gains from efficiency improvements while preserving public interest through its regulatory duties.

times awarded to private operators. Concessionaires usually pay a high rent because they are guaranteed exclusiveness and monopoly power. This pattern of concessioning commercial activities may lead to prices that are double those outside of the airport.

Individual government regulation is almost absent in this context. Being public monopolies already means interference, making it unnecessary to have economic regulation aimed at greater efficiency. Nevertheless, because of the international nature of air transport and the required coordination of activities, the ICAO has established some regulatory principles regarding airports' pricing mechanisms (discussed later) and nondiscriminatory practices due to aircraft nationality. Other rules concern the recognition of aircraft certificates and the need to facilitate customs procedures, but their main concerns are safety and security at the operator level and in air traffic control.

The Movement toward Privatization

When governments start worrying about the burden of airport financing and the lack of efficiency, the traditional model appears to be unsustainable. Nevertheless, most airports around the world could still fit this model, and only since the 1980s have things started to change. In Europe, for instance, the privatization wave has mainly taken the form of corporatization, partial divestiture, or full divestiture (only the British Airport Authority, or BAA, for this third option). Lack of public funds and underdeveloped capital markets have made a similar model difficult to apply to developing economies, such as those in Africa, Asia, and Latin America, whose privatizations have been in the form of concessions or management contracts.

If public monopolies are being turned into private monopolies, and if consumers' interests are to be protected, some regulatory provisions are required. In this sense, an important question must be kept in mind: are airport infrastructures genuine natural monopolies, or due to their multiproduct nature, should we distinguish those activities in which the exertion of monopoly power is very likely from those in which competition is feasible and desirable? This takes us to the matter of unbundling airport activities.

In the strict sense, one airport would not be subject to competition until another nearby airport began to compete for traffic.[5] If one considers, however, that the services carried out at airports are quite numerous and different in nature, perhaps some other scope exists for introducing competitive forces. This is competition for the right to serve the market.

As shown in table 3.4, most airport activities, with the exception of operational services, may be subject to competitive forces, at least in the form of competition for the market. Hence, if subcontracting takes place, any concern about the exploitation of monopoly power should mainly regard operational activities. This is why most regulatory provisions affecting airport charges concentrate on the operational side of activities. Most cases of airport pricing regulation, either discretionary or contract regulation, principally aim to control operational charges.[6]

In looking more closely at handling and commercial activities, the question arises: will the introduction of competition for the market be

5. A special case would be one airport with several terminals that are run separately.

6. An example of discretionary regulation is the one exerted over BAA airports. The literature applies the terms discretionary or commissioned regulation as synonymous.

Table 3.4. Scope of Competition in Airport Services

Tasks	*Competition for the market*	
	Feasible	*Desirable*
Operational		
Air traffic control[a]	Yes	Undetermined
Meteorological services	Undetermined	Undetermined
Telecommunication	Yes	Undetermined
Police and security	Yes	Undetermined
Fire, ambulance, and first aid	Yes	Undetermined
Runway, apron, and taxiway maintenance	Yes	Yes
Handling		
Aircraft cleaning	Yes	Yes
Provision of power and fuel	Yes	Yes
Luggage and freight loading and unloading	Yes	Yes
Processing of passengers, baggage, and freight	Yes	Yes
Commercial		
Duty-free shops	Yes	Yes
Other retailing shopping	Yes	Yes
Restaurants and bars	Yes	Yes
Leisure services	Yes	Yes
Hotel accommodation	Yes	Yes
Banks	Yes	Yes
Car rental and parking	Yes	Yes
Conference and communication facilities	Yes	Yes

a. The ATC may be subject to other forms of private participation, as discussed earlier.
Source: Authors.

sufficient to reduce monopoly power, or should some regulatory mechanism be in place? Let us assume that an airport authority concerned with maximizing profit decides to concession a given facility or service.[7] It may award the concession to one or to several competitive operators.

7. Of course, it might decide just the opposite. In such a case, the exertion of monopoly right is clear.

For instance, it may allow only one handling agent to operate the whole airport, in which case the monopoly reproduces itself, or it may allow several competing agents to serve the airport. Similarly, it may allow only one or several restaurant operators to cater the whole airport. In this way, a regulator also needs to worry about these aspects of airport operations, even if they represent only a small part of airport revenues. Table 3.5 illustrates this idea.

Once a regulator decides to fix prices, it should also be concerned with the consequences of the adopted measures. To what extent should airport quality be affected? How can it measure and control airport performance to accomplish the degree of regulation? Because these questions are essential to any regulatory framework, they are considered later in detail.

Experiences in Airport Privatization

The British began to reconsider the traditional airport model in 1987 when the government decided to take the BAA under full flotation, except for a single golden share (a share that incorporates a veto right) that was retained. When other governments chose to privatize their airports, they did not follow the same path, making the British case unique. Therefore, one should keep in mind that a great variety of privatization forms may fit into the airport infrastructure case.

One can categorize the different models of airport ownership and management as follows (Kapur 1995):

- Public ownership and public operations
- Public ownership and public operations with commercial orientation
- Regional ownership and operations

Table 3.5. Monopoly Power in Airport Handling and Commercial Activities

Handling and commercial		
Concessioned		
One operator	*Several operators*[a]	*Not concessioned*
Yes	No	Yes

a. Regulatory measures prevent collusion.
Source: Authors.

- Public ownership with private operations: joint ventures, partial/majority divestitures, management contracts, build-operate-transfer (BOT) and similar concession schemes, and so forth
- Private ownership and private operations.

Public Ownership and Public Operations. This is the model traditionally used to operate airports around the world. Usually, a civil aviation department, under the supervision of the ministry of transport or even the ministry of defense, owns and operates most airports. The Comando de Regiones Aéreas (an arm of the air force), which owned, administered, and operated a total of 400 airports in Argentina, has constituted, until recently, an extreme case of this type of model. In general, most countries begin airport service operations with the army, later distinguishing between the control and operation of military and civil air traffic services.

Public Ownership and Public Operations with Commercial Orientation. Also known as public corporations, this model attempts to improve management and airport finance autonomy, facilitating access to private capital markets. The BAA, established in 1966, was the first authority operated according to such criteria. The Israeli Airport Authority, Aeropuertos Españoles y Navegación Aérea (AENA) in Spain, and INFRAERO in Brazil are other examples of this model.

The Spanish airport model may illustrate the evolution from pure state ownership to public corporations with a bit of private sector involvement. Until 1977, the air force was responsible for providing airport and air traffic services; after that, activities were transferred to the government, with Organismo Autónomo de Aeropuertos operating airport related activities and the Civil Aviation Authority managing air traffic services. Finally, in 1990, both were merged to form AENA, a public company with autonomous status under the tutelage of the Department of Transport. Nevertheless, AENA introduced some private participation in the financing and construction of new infrastructure. For instance, it applied a build-own-operate-transfer scheme for the construction of a new terminal at Palma de Mallorca, where the selected developer was a joint venture company made up of a private promoter and AENA itself. It also has applied a similar scheme to construct a new cargo terminal at the Barcelona airport.

Amsterdam Schiphol Airport in the Netherlands is an interesting variation of this type of model. The government has a 76 percent share of participation, 22 percent belongs to the city of Amsterdam, and Rotterdam holds the remaining 2 percent. The airport follows a business-oriented approach and has financial independence, although the government may finance infrastructure investments. In spite of its public nature, the airport has managed to sell bonds in the Euromarket, getting a triple "A" rating, the highest possible bond qualification.

REGIONAL OWNERSHIP AND OPERATIONS. This is an alternative to public ownership and operation by a national body. It seeks to promote development for the airport region, putting property either in the hands of one or several local or regional entities. This approach has been used at airports in the United States (except for airports in the Washington, D.C., area),[8] the United Kingdom (except for BAA airports), and France. Some local governments may operate several airports (Aèroports de Paris has four), but the majority control just one.

At U.S. airports, despite being under local, regional, or even state supervision, many activities (up to 90 percent of total airport activity) are contracted out to the private sector. Interestingly, debt financing for the funding of infrastructure projects has been commonplace at U.S. airports. To guarantee this debt, U.S. airports used to keep long-term residual agreements with the airlines, which committed them to cover airport operating costs and debt service. The usual procedure was the following: Each year, the airport would calculate what part of the costs could not be covered by non-airline revenues, and this amount would be the required payment for air carriers. In turn, airlines would keep a great deal of operational control at the airport, including exclusive gate use and the right to approve all capital improvement programs. The changes the air transport market experienced after the passage of the Deregulation Act in 1978 reduced the value of such guarantees. Since then, airports have been shifting to compensatory agreements, which give airport authorities greater control over operations and investment plans, allowing them to charge airlines for the space used. U.S. airports may also benefit from the Airport Improvement Program, which the federal government implemented. Its funds come from taxes and user fees.

PUBLIC OWNERSHIP WITH PRIVATE OPERATIONS. Privatization policies arise in search of efficiency in a public budget constraint environment and also driven

8. In 1987, the federal government established the Metropolitan Washington Airport Authority which was given a 50-year lease to operate both Ronald Reagan National and Dulles International airports.

by disenchantment with public sector performance. Airports privatization does not have a unique model, however. The range of possible options is wide, including joint ventures, partial/majority divestitures, management contracts, BOT and similar concession schemes, and so forth.

- *Joint ventures.* One example is Kansai International Airport (Japan), which has a unique ownership structure. The Japanese government owns two-thirds of the shareholdings, with the rest belonging to 12 different local governments and more than 800 private companies and individuals. The total project cost exceeded more than US$20 billion, which included the construction of an artificial island. Under the supervision of the Ministry of Transport, the airport is administered as a private company with limited managerial and financial autonomy.
- *Partial/majority divestitures.* The government reduces its equity participation either in part or to one single share (or even to zero shares). Shares divested can be sold directly to local or regional governments or to private individuals, or they can go under public flotation. Divestitures are mainly used as a means of obtaining private equity funding for future airport expansion. The only instance of majority divestment is the BAA, as mentioned earlier. Instances of partial divestitures are the Zurich, Vienna, and Copenhagen airports. The Zurich airport (Switzerland) is an interesting case because, although the Canton of Zurich retains property, the airport is operated by a private company (Flughafen Immobilien Gessellschaft), which in turn belongs to the canton (with 50 percent of the shares) and a group of private individuals. The Vienna airport (Austria), originally a public corporation, is today, after a partial divestiture, 48 percent in public hands, including the participation of Amsterdam Schiphol Airport. After the BAA, it was the second airport quoted at the stock exchange. The Copenhagen airport put 25 percent of shares under flotation in 1994.
- *Management contracts.* The management of all or part of the airport is contracted out to a specialized operator for a given period of time, with certain performance, maintenance, incentive, and infrastructure investment conditions. For instance, the government of Cameroon created Aèroports du Cameroon to operate 7 of the country's 14 airports for a 15-year period. Participants in this company include Aèroports de Paris, with 34 percent of shares, and the Cameroon government, with 24 percent. The remaining shares are distributed among

carriers and a major bank. Aèroports du Cameroon is required to reinvest part of its profits, and it can establish airport charges after consulting the government and airport users.

- *Concession contracts and variants.* Perhaps the most recent and innovative arrangement, these allow airports to benefit from private sector involvement. The Argentine government provides a recent example; in February 1998, it subscribed to a concession contract with a consortium, Aeropuertos Argentina 2000, regarding a set of 33 airports that were all awarded to the same concessionaire. It established a 30-year concession period, with a possible 10-year extension included in the contract. Aeropuertos Argentina 2000 has the right to collect some aeronautical charges, which are subject to economic regulation and were initially established for a five-year period.[9] Nonaeronautical charges can be set freely. The corresponding total annual payment to the Argentine government exceeds US$171 million, an amount that is periodically adjusted according to the Producer Price Index. In addition, the consortium is required to invest a minimum of US$2.1 billion. The group has already taken control of Buenos Aires' two airports, Eisesa and Aeroparque. The regulatory body specially created at the time is the Organismo Regulador del Sistema Nacional de Aeropuertos. Among other tasks, it supervises airport fees and the fulfillment of investment requirements.

  Australia provides another interesting example of an airport concession.[10] Twenty-two of its airports, which the Federal Airport Corporation[11] previously controlled, have been or are currently being leased for 50-year terms, with an option for another 49 years. According to the government, each airport should be sold separately and remain subject to a regulatory framework whenever possible. Deciding against the use of an industry-specific regulator, the government made the Australian Competition and Consumer Commission responsible for regulatory duties.

  - *BOT schemes.* This is when the government grants a concession or franchise to a private firm in order to finance and build or

9. Comando de Regiones Aéreas, which is in charge of air traffic control, controls other aeronautical charges.

10. To date, privatization of airports in the Asia Pacific region is pretty much an unknown apart from Australia.

11. Established in 1988 as a government business enterprise.

modernize a facility that will also be operated by the firm for a certain period of time (20 to 50 years is a common period for airports). The private operator will get the corresponding revenues and in turn will assume all commercial risk. When the concession period expires, the facility will return to the government. The concession contract may include some regulatory provisions regarding the prices charged or the quality provided. This scheme and all of its variants have been used widely for infrastructure development. For example, the Colombian government used a BOT scheme in 1995 for the construction and maintenance of a second runway as well as for the maintenance of an existing runway at El Dorado Airport in Bogotá. (Colombian Civil Aviation continues to provide air traffic control.) The landing fee revenues collected during the 20-year concession period will recover the US$100 million of project costs. In this case, the government granted a minimum level of revenue and assumed a great part of the risk. This may represent a difficult blueprint barrier, however, for future privatization projects in Colombia. Indeed, it seems that plans to concession the Cali airport failed because bidders were expecting the same commercial risk protection.

- *Build-own-operate-transfer schemes.* Slightly different, this system allows the private operator to also retain ownership of the facility during the concession period, usually to guarantee bank loans. Development of Toronto's Lester B. Pearson Airport's third terminal, with a capacity for 10 to 12 million passengers, was under this type of arrangement. The deal included a 40-year land lease, with an option to renew for an additional 20 years, a CAN$30 million lump sum payment to the government, and an annual lease payment based on developers' gross revenues. Toronto's airport represented a rare combination of public and private ownership and operation. The government body Transport Canada owned and operated terminals one and two, while terminal three used to be privately owned, but Lockheed Air Terminal of Canada Inc. operated it under a management contract. Transport Canada coordinated activities, provided air traffic control, and was the proprietor of runways and taxiways. Charges at terminal three were twice as high as those at other terminals, segmenting the market: the more prestigious international carriers tended to use terminal three, while low-cost and/or regional carriers mainly used the other terminals. At the moment, Toronto's third terminal is again under public sector management.

- *Lease-develop–operate scheme.* This is another alternative for introducing private participation in airports. It consists of a long-term concession on an existing facility. A private firm operates and upgrades or expands the facility, obtains revenues from operations, and pays rents to the government, which retains the property throughout the concession period. This type of arrangement was planned for La Chinita Airport in Maracaibo (Venezuela) in 1993, but it was unsuccessful due to a consortium breach of contract and changes in the political situation.

Private Ownership and Private Operations. This is exemplified by airports the BAA operates in the United Kingdom.[12] The BAA used to be a public corporation until 1987, when the government, applying the Airports Act, decided to take 500 million shares under full flotation at a subscription price of £2.40 each. As mentioned earlier, the government kept a single share (golden share), and reserved 25 percent of equity for employees. To avoid capital concentration, it limited individual participation to 15 percent. It initially limited foreign capital participation, although it reaches 10 percent. Private participation amounts to 95 percent of total shareholdings. The Airports Act also provided for the regulation of the BAA to avoid any exploitation of monopoly power. The government appointed the Civil Aviation Authority as regulator, also giving the Monopolies and Mergers Commission (MMC) and the Office of Fair Trading the ability to review BAA activities.

Another full divestiture example is Belfast International Airport, although a public tender mechanism was selected. A group of managers and employees presented the winning US$72 million bid. In contrast to the BAA, it is not subject to price regulations.

Also worth mentioning is that occasionally a private sector company builds and operates an airport by itself. An example is London City Airport.

Conclusion. As we have seen, a wide range of possibilities exists for private sector involvement in airports (see tables 3.6 and 3.7). An ideal model

---

12. The BAA manages Heathrow, Gatwick, Stansted, Glasgow, Edinburgh, Aberdeen, and Southhampton airports.

Table 3.6. Inventory of Airport Ownership Structures in Selected Countries

	Public participation			Private participation	
	Public risk			Shared risk	Private risk
Region	Government department	Public corporation	Regional government ownership	Joint public-private venture	Private ownership
Europe and North America	Czech Republic Greece Hungary Romania Russia Sweden	Austria Canada Germany Ireland Israel Netherlands Norway Spain	United Kingdom France United States	• **Share flotation** Brussels (Belgium) Liverpool (UK) East Midlands (UK) Copenhagen (Denmark) Italy Vienna (Austria) Zurich (Switzerland) • **BOT** Birmingham (UK) • **BOOT** Toronto (T3) (Canada)	• **Share flotation** BAA (UK) • **BBO** London City (UK) • **MEBO** Belfast (UK)
Asia and Pacific	China Hong Kong (China) Malaysia India Japan Thailand	New Zealand Singapore	None	• **Joint venture** Kansai (Japan) • **Concession contract** Australia	• **None**

(table continues on following page)

Table 3.6 continued

	Public participation			Private participation	
	Public risk			Shared risk	Private risk
Region	*Government department*	*Public corporation*	*Regional government ownership*	*Joint public-private venture*	*Private ownership*
Latin America and the Caribbean	Venezuela Haiti	Jamaica Brazil	None	• **LDO** Maracaibo (Venezuela) • **Concession contract** Argentina • **BOT** Mexico City Bogota (Colombia)	• **BOO** Punta Cana (Dominican Republic) Freeport (Bahamas)
Middle East and Africa	Angola Gabon Kenya Saudi Arabia	Nigeria South Africa	None	• **BOT** Istanbul (Turkey) • **Management contract** Cameroon	• **None**
Financing sources	• Direct government subsidies • Multilateral lending • Bilateral lending	• Debt with government guarantees • Municipal bonds		• **Debt with government guarantess** • **BOT, BTO, leases** • **Quasi-equity instruments** • **Equity instruments**	

Source: Adapted from Kapur (1995).

Table 3.7. Structure of Some European Airports

Country	*Type of ownership*	*Market structure*	*Pricing principles/subsidies*
Spain	**Airport:** Publicly owned (AENA–state-owned national airport authority)	**Handling:** Monopoly for third-party passenger handling **Operating services:** Monopoly	Landing fee based on weight, different passenger fees based on destination. Yearly regulation, no discrimination
	ATC: Publicly owned by AENA	Monopoly	Based on aircraft type and distance flown over own airspace
France	**Airport:** Publicly owned (Paris airports are owned and operated by Aéroports des Paris, which is owned by the state)	**Handling:** Self-handling is allowed, but not third party **Operating services:** Monopoly	Aeronautical charges regulated by the state. Landing fees based on weight. Passengers charged on departure
	ATC: Publicly owned	Monopoly	Based on aircraft type and distance flown over own airspace
Germany	**Airport:** Publicly owned except Dusseldorf and Berlin airports	**Handling:** Monopoly/oligopoly **Operating services:** Monopoly	Takeoff and landing fees are regulated by air transport authorities
	ATC: Publicly owned	Monopoly	Based on aircraft type and distance flown over own airspace
Sweden	**Airport:** Publicly owned (state), municipality/mixed and one private. (Swedish Civil Aviation Administration runs all major airports)	**Handling:** Monopoly. Except for SAS and passengers, not allowed **Operating services:** Monopoly	Charges regulated by state. Landing, terminal navigation, and security charges. Landing fee based on weight (different rates for international/domestic flights)
	ATC: Publicly owned	Monopoly	Based on aircraft type and distance flown over own airspace
Switzerland	**Airport:** Mixed ownership between a public agency and private	**Handling:** Monopoly (by Swissair) **Operating services:** Monopoly (Swissair)	Charges for landing (combined with noise charge), passengers, aircraft parking)
	ATC: Publicly/Privately owned	Monopoly. Nonprofit company	Based on aircraft type and distance flown over own airspace

(table continues on following page)

Table 3.7 continued

Country	*Type of ownership*	*Market structure*		*Pricing principles/subsidies*
Netherlands	**Airport:** Publicly owned	**Handling:** Competitive. Three ground handling operators	**Operating services:** Monopoly	No subsidy for international; regional services are subsidized in 60 percent of their total operating costs
	ATC: Publicly owned	Monopoly		Based on aircraft type and distance flown over own airspace
Italy	**Airport:** Publicly owned except Rome and Naples airports	**Handling:** Monopoly (undergoing liberalization)	**Operating services:** Monopoly	Subsidies/market. Subsidies mainly for operations and infrastructures
	ATC: Publicly owned	Monopoly		Based on aircraft type and distance flown over own airspace
Belgium	**Airport:** Publicly owned	**Handling:** Three ground handling operators	**Operating services:** Monopoly (local)	Landing, passenger, parking, air bridge, and airport fuel fees
	ATC: Publicly owned	Monopoly		Based on aircraft type and distance flown over own airspace
Portugal	**Airport:** Publicly owned by ANA. Privatization plans for ANA announced	**Handling:** Passenger – monopoly (TAP). Freight – some competition	**Operating services:** Monopoly (ANA)	Aeronautical charges set by government. Landing fee based on weight; no discount or surcharge for noise
	ATC: Publicly owned	Monopoly		Based on aircraft type and distance flown over airspace
United Kingdom	**Airport:** Most major airports privately owned by one company (BAA), with Manchester and smaller airports owned by local authorities.	**Handling:** Competition	**Operating services:** Monopoly, but commercially run	BAA and Manchester: RPI-X constrained, fixed charge per aircraft (including surcharge of 50 percent for noise) and a per passenger tax, with surcharges for parking
	ATC: Publicly owned (plans announced to privatize)	Monopoly, but commercially run		Based on aircraft type and distance flown over own airspace

Source: Viegas and Fernández (1997).

has not emerged. The BAA case provides enough evidence to conclude that full divestiture allows an improvement in market efficiency. Poole (1990) reports that the number of passengers handled per employee increased after privatization, while at the same time operating expenses declined. Nevertheless, the procedure used to privatize the BAA may not always be applicable. First of all, it requires developed capital markets, which is quite rare in developing economies. It also needs a new regulatory framework, which is costly and difficult to implement. In addition, it is not a feasible option for governments that wish to retain property for political reasons. These are the main reasons undermining the appearance of alternative privatization procedures. Nevertheless, a dominant model that falls in the middle of the privatization spectrum seems to be emerging in Latin American countries, as shown in table 3.8. This is the concession model in any of its variations. It seems to provide governments with much-needed funds for airport infrastructure expansion. At the same time, it allows them to keep the property and retain the facilities at the end of the concession period. Furthermore, it provides a financial windfall for governments with restricted budgets.

Any concession process is , however, complex and costly. From the initiation of economic and technical studies to the time when the concession contract is ready, the whole process may take several years. In addition, transparency is essential when awarding private concessions. Otherwise, political corruption or lawsuits may be the final outcome.

Price Regulation

The trend toward privatizing the airport industry stems from government views that airports ought to be financially self-sufficient. Some regulatory provisions, however, must be in place to control the substantial monopoly power that airports possess. For example, in the case of privatized British airports, the MMC (1996) reported that in certain cases, airports in London had observed a course of conduct that was against the public interest. As Forsyth (1984) asks, can regulation be a means of improving airport efficiency through limiting monopoly power, and how will regulation influence the equilibrium between productive and allocative efficiency? The answers to these questions depend on the features of the airport industry and its regulatory system.

We must clearly distinguish among airport activities to determine those that can be monopolized. The classification that separates aeronautical and nonaeronautical services, as discussed earlier, is adequate for our purposes. A great variety of commercial activities carried out at an airport, such as

Table 3.8. *Privatization Processes in Latin American Airports*

Country	*Privatization plans*
Argentina	Privatization in progress • 33 airports concessioned as a group to Argentina 2000. • Concession for 30 years with a possible 10-year extension. • Bidding variable: annual payment. • Total investment: Approximately US$2,000 million. • Excluded: ramp services, cargo, and duty-free shops. • ATC: Fuerza Aérea Argentina.
Bolivia	The three largest airports have been already privatized • El Alto, Viru Viru, and Cochabamba concessioned to Airport Group International. • Concession for 25 years, beginning March 1997. • Bidding variable: percent of revenue—14 percent minimum. • A fund was created for maintenance and operation of the 34 remaining airports. • Adaptation to FAA II rules and IATA level B.
Brazil	Strategy under consideration.
Chile	Privatization in progress • Concepción, Punta Arena, Temuco, and Copiacó to be concessioned. • Investment requirements: US$150 million. • Concession for 15 years. • Bidding variable: lowest charge per epax. Minimum revenue guaranteed. • Excluded: aircraft fuel services. • ATC: DGAC. Airports already concessioned: Iquique, Calama, La Serena, Puerto Montt, and Santiago.
Colombia	Bogotá: El Dorado second runway concession to Ogden-Dragados-Conconcreto. Cartagena: awarded to Schiphol (30 percent) for 15 years. Fixed annual payment US$24.5 million. Barranquilla: awarded to AENA (50 percent) for 15 years. Fixed annual payment US$9 million. Medellin and Cali: next in line.
Costa Rica	Privatization in progress. OD contract for San Jose International Airport is being prepared.
Dominican Republic	Privatization in progress. OD contract for Las Américas, Puerto Plata, Samana, and Barahona.
Ecuador	Privatization in progress. BOOT contract for new airport development at Quito and Guayaquil. Required investment of US$700 million.
El Salvador	Privatization under study.
Guatemala	Privatization under study for La Aurora and Tikal.
Honduras	Privatization under study for Tegucigalpa, San Pedro Sula, La Ceiba, and Roatan.
Jamaica	Privatization for Montego Bay-Sangster • BOO for passenger terminal. • 49-year term. • The concessionaire will also operate actual terminal and airside activities.
Panama	Privatization under study.
Peru	Privatization plans for five national airports in the first half of 1999 under a master concession.
Mexico	Privatization in initial stage. • 58 airports to be concessioned, grouped in three sets. Mexico D.F. excluded. • Southeast Airport Group (Cancun): awarded to the consortium formed by Copenhagen airport, GTM, Cintra, and Tribasa.
Uruguay	Concessions plans: • Laguna del Sauce and Punta del Este. • Carrasco and Montevideo
Venezuela	Privatization under study for Simon Bolivar airport at Caracas.

Source: Adapted from Anuario del Transporte (1997).

tax-free shops, retail shopping, restaurants, hotels, and bank services, are considered nonaeronautical. For these types of activities, introducing competition would be feasible and desirable. Likewise, the unbundling of activities could be useful in reducing the exertion of monopoly power for a small set of aeronautical services related to aircraft movement, such as the provision of runways, aprons, and taxiways. Therefore, if an airport is to be privatized, the establishment of controlling rules that regulate private sector involvement is clearly needed. Regulation could take several forms, with the main one being competition for the right to serve the market (concessions or leasing) as well as fare and/or profit controls. The most common regulation tool used for limiting monopoly power, however, is price regulation. Some of the particulars of airport pricing structures need to be pointed out before these price control mechanisms can be explained.

An airport pricing system has to deal with several features, including cost coverage, congestion, environmental impacts, standard level of services, investment plans, and cross-subsidies. Treating each of these alone is complex. Even more difficult is conciliating all of these elements under a common pricing policy. For example, the financial goal of cost coverage must be in accordance with the necessity of investing in additional capacity. The pricing structure not only must ensure the allocative efficiency of actual resources, but it must also reflect the need for new capacity and its efficient assignment. Hence, one must determine the optimal level of capacity (and therefore of congestion) at the airport. We should also add that the multiproduct nature of airport activities implies the presence of joint costs that are common to the operation of several services. For instance, common areas in the terminal buildings allow the processing of passengers (handling), while at the same time these areas are also used for commercial purposes. This makes determining the correct cost allocation for different airport services difficult. Furthermore, the airport industry shows increasing returns to scale for aeronautical operations, due to capital investment indivisibilities. These characteristics clearly influence the airport pricing structure. How could these peculiarities be incorporated into the pricing structure and connected to the design of a regulatory framework?

Traditional Pricing Policy

International organizations such as the ICAO and the IATA have recommended that airport cost coverage include the application of average costs as the basic price. In addition, these organizations have sought to establish a uniform fare structure for the whole industry. Dividing incurred costs by the number of processed traffic units provides a unitary tariff. This procedure could provide several fares for each service by distinguishing among the different components of total cost. Given that all users pay the same amount for the use of the same services, most airlines support this mechanism as objective and fair. The reality, however, is that different operators impose different costs and therefore should face different charges. For example, an airline that operates during peak periods imposes a capacity cost that is higher than those operating during off-peak periods. Finding a way to incorporate this and other industry particularities into the actual fare system is needed within the context of regulation. Otherwise, we have to consider alternative pricing mechanisms.

Fare structures found at the majority of airports are similar because most countries follow ICAO and IATA guidelines. Both organizations seek a uniform pricing system, recommending the use of aircraft weight as the basis for the estimation of applicable charges. Table 3.9 shows the

Table 3.9. *Airport Charges at Selected Airports, 1998*

Charges	*Rio de Janeiro*	*Manchester*	*Sydney*	*Madrid*
Landing fee				
Basic unit	MTOW[a]	MAW[b]	MTOW	MTOW
Charge per	Ton	Ton	Ton	Ton
Increases with weight	No	No	No	Yes
Free parking	3 hours	4 hours	2 hours	3 hours
Surcharges/rebates				
Night lighting	No	No	No	No
Noise	No	Yes	Yes[c]	No
Passenger charges				
Paid by	Passenger	Company	[c]	Company
Distance related	No	No	No	No
Other charges				
Security	No	Yes	Yes	Yes
Rescue/fire service	No	No	Yes	No
Airbridge	No	No	No	Yes
Terminal (general)	No	No	No	No

a. MTOW: Maximum takeoff weight.
b. MAW: Maximum aircraft weight.
c. Not an airport charge. It is collected at ticketing point as government levy.
Source: Doganis (1992), adapted and actualized.

pricing structure of several countries, verifying that such a structure basically corresponds to a landing fee calculated according to aircraft weight, plus a departure fee for passengers.

The private sector's increasing involvement in airport activities might break the uniformity of pricing structures around the world, leading to a more efficient pricing system at privatized airports. For a private firm, coverage of actual costs—as well as coverage of those costs generated by future investments in additional capacity—is of critical importance. The actual pricing structure upon which regulatory devices are applied must be consistent with additional capacity investment so that corresponding costs are also covered. Because the allotted period to recover the investment is long, the regulator should permit price variations during the investment period with the aim of adjusting costs and generating revenue. Among the various problems that a regulator might encounter, however, are the difficulty of establishing credible commitments and the need to develop a deep knowledge of the operations and opportunities of a privatized airport.

The selection of the initial price structure will be the basis for applying the regulatory mechanism. It should be an adequate guideline for future investment and also ensure the efficient allocation of resources. Economic theory states that if the price is established according to the service marginal cost, an efficient allocation of resources among users is obtained. The paid fare reflects the true service value, and those who are not willing to pay are not served. Those airports that generally operate below available capacity, however, present a very small marginal cost and hence will not produce enough revenue to cover total costs. In the airport industry, many costs are sunk, or historical costs exist that do not conform to the service marginal cost. Therefore, strictly applying a charging policy following the marginal cost criterion would inevitably lead to financial losses for those airports that operate below available capacity.

Because the price-demand elasticity of airport services is less than one, as discussed earlier, another possibility for generating extra revenue might be through applying an ad hoc rule known as Ramsey pricing. This policy suggests that when the marginal cost rule does not allow enough revenue generation to cover costs, charging users according to their willingness to pay is more efficient. Then cost coverage is assured without getting far away from the efficient allocation principle. Hence, this would be a means for deficit reduction that avoids the use of cross-subsidies. Airport monopoly power, however, would be substantially exploited.

The British Airports: Price Regulation Through an RPI-X Formula

The BAA now enjoys a considerable degree of market power. The majority of air traffic arriving or departing the United Kingdom goes through two of the most important BAA airports, Heathrow and Gatwick. The chance for competition from other airports in the United Kingdom and on the European continent, such as Paris or Amsterdam, is remote. The possible appearance of a competitor would be frustrated by the occasional adequate fare cut at London airports. Hence, the monopoly power BAA airports exert is real, and it may have repercussions on service, users, and society as a whole.

The Civil Aviation Authority (CAA) of the United Kingdom is responsible for providing air traffic control services and regulating safety and economic aspects at the country's airports. Among its objectives as an airport regulator are the protection of consumer interest, the promotion of economic efficiency, the financial viability of airport services, and the encouragement of additional capacity investments to meet future growth in air transport demand. The CAA is most known, however, for establishing a maximum level of charges for large airports. The Airports Act (1986) does not specify anything regarding the regulation of the BAA's commercial activities. The only charges subject to regulation are landing, passenger, and aircraft parking fees. Profits generated by commercial activities are usually used to compensate for low, regulated aeronautical fares. Therefore, aeronautical services have a cross-subsidy with revenues arising from commercial activities. Such a mechanism is known as the single till principle. Obviously, applying this principle leaves aeronautical service prices below provision costs, which generally represents a problem in the case of a congested airport. Consequently, applying this method leads to economic inefficiency. Nevertheless, to abandon it would imply that aeronautical service charges should reflect the higher provision costs, which would lead to airports increasing their profits because they would no longer need to cross-subsidize and could therefore make larger profits in nonregulated commercial activities. Under the single till principle, air carriers also enjoy part of airport commercial revenues and cross-subsidizing keeps aeronautical charges reasonable. It also ensures that the private airport operator would not obtain excessively high profits. This is the rationale behind the behavior of the British airports' regulatory authority. Of course, applying the principle does not help solve airport congestion problems.

Price regulation takes the form of a price cap applied to revenues deriving from airport charges per passenger, also called revenue yield (see box 3.3). Price cap regulation according to the RPI-X formula has been a key element in

Box 3.3.* *The RPI-X Formula for BAA Airports

Regulation of fares through an RPI-X mechanism applied to revenues from airport charges (landing, passenger, and aircraft parking fees) implies that revenue per passenger should not exceed a given maximum value determined by the following expression:

$$M_t = [1 + (RPI_t - X_t)/100]\, Y_{t-1} - K_t$$

where M_t = maximum allowable revenue per passenger for year t; RPI_t = percentage of change for the Retail Price Index between years t and $t-1$; X_t = factor X (percent) in year t; and Y_{t-1} = revenue per passenger in the year $t-1$, calculated according to the following formula:

$$Y_{t-1} = [1 + (RPI_{t-1} - X_{t-1})/100]\, Y_{t-2} + S_{t-1}$$

where S_{t-1} = the allowable security cost per passenger in the year $t-1$ (it corresponds to 95 percent of the annual equivalent); and K_t = correction factor per passenger applied in year t (whether of a positive or negative value), which can be obtained through the formula

$$K_t = [1 + I/100]^2\, [T_{t-2} - (Q_{t-2} \cdot M_{t-2})]/Q_{t-2}$$

where T_{t-2} = total revenue from airport charges in year $t-2$; Q_{t-2} = passenger volume in year $t-2$; M_{t-2} = maximum allowable revenue per passenger for year $t-2$; and

$$I: \text{if } K_t > 0 \Rightarrow I = SR + 3 \text{ percent}$$
$$\text{if } K_t < 0 \Rightarrow I = SR.$$

SR (specified rate) is the average of the discount rate for public funds (expressed as a percentage). The Bank of England publishes this value weekly for a 12-month period starting at the beginning of October of year $t-2$ through the end of September year $t-1$.

the field of regulatory reform in the United Kingdom, where approximately 50 firms are under this sort of regulation. This system consists of a pricing structure subject to specified maximum fare increases, expressed in terms of percentages that cannot exceed the difference between the Retail Price Index and a given factor X. This index is preferred to an industry-specific one because the regulated firm cannot manipulate it. After an established period (usually five years), prices and limits are revised. The *X* factor, which is exogenous to the firm, may vary for each year of the regulatory period.

Notably, the number of processed passengers is not the only output at airports. This type of regulatory system (revenue yield) does not consider aircraft that carry cargo and mail. An alternative regulatory application is the tariff basket approach, in which the regulatory mechanism is applied to a weighted average of each component of the fare structure. This approach takes different airport outputs into account by weighting each element of the fare structure based on the revenue it generates. The British CAA, however, recommends regulation based on passenger revenue. No evidence yet points to the existence of serious problems in applying this method.

Applying a price cap formula may also allow part of the costs to pass directly to users. For instance, at the BAA's London and Manchester airports, 95 percent of the additional security costs the Ministry of Transport imposes are permitted to pass through, with a one-year lag period. The regulator may opt to allow a high price to compensate for the risk of losses, or it may reduce the period of regulation as a means of minimizing risk. This last alternative aims to protect airports against unexpected cost changes. Table 3.10 is an example of applying the price cap formula at the Manchester airport for a five-year period.

As the table shows, the formula is adjusted to allow 95 percent of security costs to pass through to users. A correction factor based on passenger traffic forecasts also permits the adjustment of forecasting errors that might lead to differences between allowable and actual revenues obtained. Box 3.4 explains in detail the calculations and terms used to construct table 3.10.

Table 3.10. *Application of a Price Cap at Manchester Airport, 1993/94–1997/98*

Variable	*1993/94*	*1994/95*	*1995/96*	*1996/97*	*1997/98*
X (percent)	3.0	3.0	3.0	3.0	3.0
RPI (percent)	1.8	2.2	3.9	2.1	3.5
RPI – X	–1.2	–0.8	0.9	–0.9	0.5
£ per passenger based on RPI – 3 (M_t without including S_t and K_t)	7.675	7.614	7.683	7.614	7.652
Security costs adjustments (S_t in £)	n.a.	n.a.	n.a.	0.172	0.173
Correction factor (K_t in £)	n.a.	n.a.	0.265	0.379	0.911
Maximum allowable revenue per passenger (M_t in £)	7.675	7.614	7.948	8.165	8.736[a]
Revenue obtained per passenger (£)	7.435	7.278	7.136	7.192	7.505[a]
Difference	–0.240	–0.336	–0.812	–0.973	–1.231[a]
Revenue losses (million £)	3.1	4.8	12.0	14.2	19.0

n.a. Not applicable.
a. Estimated values.
Source: Monopolies and Mergers Commission (1997).

Box 3.4 Details of Applying the Price Cap at Manchester Airport

Applying the formulas in box 3.3 permits the generation of the values in table 3.10. Figures were obtained starting with a base year reference value of revenue per passenger of £ 7.768. According to the formula, for correction factor K_t, this only makes sense from year three onward. The period considered goes from 1993/94 to 1997/98.

Year 1993/94
$Y_{t-1} = 7.768$
$K_t = 0$
$M_{93/94} = [1 + (1.8 - 3.0)/100]\ 7.768 = (0.988)\ (7.768) = 7.675$

Year 1994/95
$Y_{t-1} = 7.675$
$K_t = 0$
$M_{94/95} = [1 + (2.2 - 3.0)/100]\ 7.675 = 7.614$

Year 1995/96
$Y_{t-1} = 7.614$
$K_t = 0.265$
$M_{95/96} = [1 + (3.9 - 3.0)/100]\ 7.614 = 7.683$

Year 1996/97
$Y_{t-1} = 7.683$
$K_t = 0.379$
$S_{t-1} = 0.172$
$M_{96/97} = [1 + (2.1 - 3.0)/100]\ 7.683 = 7.614$

Year 1997/98
$Y_{t-1} = 7.614$
$K_t = 0.911$
$S_{t-1} = 0.173$
$M_{97/98} = [1 + (3.5 - 3.0)/100]\ 7.614 = 7.652$

Note: M_t values do not include S_t and K_t. It corresponds to the fourth row of table 3.10.

When limits on prices are imposed, profitability may be increased at the expense of service quality. For instance, an airport may reduce costs by not cleaning the terminal building regularly or by allowing congestion and delays. Hence, when prices are regulated through a price cap, monitoring quality by establishing reasonable standards is always necessary. This was a crucial element when airlines evaluated the quality of service at BAA airports. Carriers argued that the absence of standards might be an incentive for the BAA to increase profits by reducing the quality of service.

A regulator must also consider that airports may try to cross-subsidize aeronautical activities when subject to regulation. With the presence of joint costs, they are tempted to allocate a great part of these to the regulated activity, or to charge monopolistic prices for unregulated commercial services in which price control is more difficult. In this sense, the BAA has argued that cross-subsidization from commercial services was necessary, because as a result of the strict control, regulated aeronautical fares were quite low. The main consequences of this were the diversification of provided services and an emphasis on commercial activities.

Another element to take into account is that an efficient fare structure requires great flexibility in its application, due to the changing nature of demand for airport services. According to the BAA, price controls clearly affected the efficiency of its services. The BAA also asserts that severe regulation may result in financial difficulties for the airport operator, bringing unforeseeable consequences for profits.

Finally, according to the CAA, the main benefit derived from regulation was that it obliged airports to keep costs low. In other words, airports were minimizing costs to reap higher profits. Other important conclusions were (a) the regulator must clearly know what its goals and responsibilities are; and (b) the regulator must have direct access to all the necessary information, including confidential material, to carry out its work properly.

The Problem of Airport Costs

As we have pointed out, some aspects of the airport industry are difficult to incorporate into the regulatory structure. Nevertheless, if regulation through an RPI-X formula is to be efficiently applied, these elements must be taken into account. Factors that may cause greater problems when regulating airport charges are, among others, congestion, externalities such as noise (see earlier section), investment indivisibilities, and quality of service (see Forsyth 1997 for more on the problem of airport costs).

CONGESTION. The costs of processing an additional passenger or aircraft at an airport that operates below available capacity at any time are close to zero.[13] Under these conditions, additional passenger or aircraft charges should be established according to the airport's short-run marginal cost. If demand increases, however, causing a large concentration of traffic during peak hours, the corresponding marginal cost would be much higher than the one applicable during off-peak periods. In this case, price discrimination would be

13. This section assumes a price cap directly applied on airport fares.

justified because the price charged for peak periods could be much higher than the one applied to off-peak intervals. If the goal of investment is to increase capacity, fares must also incorporate this goal. In summary, an optimal fare structure that accounts for congestion problems needs to be flexible. If price controls take the form of price caps, however, rigidities will not permit changes in prices over time. In practice, this regulatory mechanism limits the use of prices as a tool for managing the problem of congested airports.

Two important aspects relate to congestion at airports: (a) determining optimal capacity and (b) determining its efficient allocation. Regarding the former, the existence of a price cap implies that the airport has no incentive to optimize the available capacity because it faces a fixed fare structure for which revenues increase only if traffic flow also increases. This type of price regulation breaks the link between congestion reduction and revenue generation. In other words, the airport does not gain from reducing congestion. A possible solution may be incorporating congestion costs into the price regulatory formula. Nevertheless, finding an adequate indicator of congestion is not easy, and including a variable for it in the price cap would be even more difficult. A possible way around this is a regulator that establishes the optimal level of capacity as the result of a cost-benefit analysis that compares congestion costs to the benefits that arise from a larger available capacity.

Once optimal capacity has been determined, it must be efficiently allocated. This usually consists of determining a price that equilibrates market supply and demand. Those airports with traffic volumes exceeding capacity at certain times should apply different charges at peak and off-peak periods. The high level of traffic during peak periods justifies price discrimination. A high enough level may lead to a need for additional capacity investments. Yet the above-mentioned conflict still remains. Price cap regulation limits this possibility because its goal is to keep fares low. This is incompatible with peak pricing because peak fare is necessarily higher to allocate capacity more efficiently.

A possible way to reconcile the application of a price cap regulatory formula with the efficient allocation of capacity at congested airports consists of applying the price cap to average fares or average revenues, as in the BAA case. In addition to the price formula, a mechanism for allocating slots and/or establishing a slot market may be applied. For instance, available capacity as determined by the regulator could be allocated through a public auction, after which resale would be permitted. The main problem with this procedure is determining who obtains the rents from the sales. If the regulator allows the airport to take the money, the airport has an incentive to keep capacity scarce and prices high. This experience, however, seems to have worked relatively well at airports in London.

EXTERNALITIES. Noise is one of the most important negative externalities generated at airports. Because aircraft noise affects a large number of people, the internalization and incorporation of its effects into total airport costs is needed. In order to proceed, one must estimate the external marginal costs and then establish a fare structure. The main question, however, is how to jointly consider the external effects and the regulatory framework. In this sense, two main problems exist: first, how to incorporate noise control devices, for instance (through a special fare to mitigate excess noise into the regulatory formula); and second, how to reconcile permissible noise levels and airport capacity.

In general, three alternatives for regulating noise level are consistent with the RPI-X formula: (a) incorporating a noise index into the formula, (b) charging a special fare paid by the airport or its users, and (c) establishing quantitative limits. The first procedure would allow airports to charge higher fares for lower noise levels, in a way that airlines would be penalized for succeeding to reduce noise. Hence, airlines would have the incentive to collude and operate in the opposite direction. The second entails airports being penalized according to the noise generated by their customers. Given that it is not the airport itself that generates noise, however, but its users, the airport should be in the position to pass on these costs to the users. An alternative is to directly charge air transport carriers. Finally, the establishment of quantitative limits (alternative three) involves restricting certain types of airplanes or banning air traffic operations at certain times of the day. This may be complemented by a charge aimed at reducing noise during peak hours. For example, night restrictions might be complemented by another charge that would limit noise during the day. Such a combination can be found at the Sydney airport, where a noise charge is combined with the application of quantitative limits.

Capacity may be increased by choosing different aircraft approach routes, which also leads to increased noise levels. One can study this trade-off through a cost-benefit analysis. The regulator needs information regarding the costs of noise for different routes and to compare it to the benefits arising from the availability of additional capacity. The regulator is then in the position to select the most efficient combination. This is only possible, however, if the regulator controls other airport aspects, including environmental impact.

QUALITY OF SERVICE Quality of service is an important aspect that must be controlled when implementing price regulation. An airport that faces a regulated price will try to reduce its costs to obtain a higher profit margin. Hence, the regulator must closely supervise elements related to quality of service. Four mechanisms control quality: First, the regulatory agency can ask the airport to publish certain quality standards. Second, a quality index can be incorporated into the RPI-X formula. A third option consists of compensating users for low-quality services. Finally, a fourth possibility is fixing minimum quality standards. Airports that do not comply are fined or subject to a revision of regulatory conditions.

Ad hoc methods are usually applied to control quality. For instance, in the British and Australian telecommunications industries, the regulator collects information through quality indexes. Those firms with quality indicators below required levels are subject to regulatory pressure. Developing good quality indicators for airports, however, is not easy. Nevertheless, within the context of regulation, taking steps to evaluate quality of service and ensure that it does not deteriorate is crucial. Fixing minimum quality standards and enforcing compliance may be the most effective means, because this implicates airports in the attainment of quality. As the main users of airports, however, air carriers also play a large role in airport quality, and they frequently work jointly with airports to provide services. Concessionaires of airport services such as passenger and luggage handling in many cases are the airlines themselves or other outside companies. Consequently, attaining quality of service standards must be the responsibility of both the airports and their main operators.

Another quality-related aspect is the existence of enough airport capacity to offer services at an acceptable level of quality. As previously mentioned, incentive is lacking for investing in new capacity at airports subject to price regulation. Uncertainty about the coverage of additional capacity costs leads to the belief that certain adjustments should be allowed in order to charge higher prices when investment takes place. This means, however, that the regulator must use ad hoc solutions that move away from the simplicity of the single application of a price cap.

INVESTMENTS. The provision of airport infrastructure is subject to the existence of significant indivisibilities, meaning that capacity can only be augmented by adding large, indivisible units. In this context, an important relationship exists between airport charges and the need to amplify capacity, which is an additional problem for the regulatory framework.

When an airport disposes of excess capacity, the optimum price is given by the short-run marginal cost. If demand increases, the use of capacity needs to be rationalized through a significant price increase, which can then be equal to the long-run marginal cost. This is the efficient way to proceed when capacity is scarce. In other words, users demanding more capacity pay the marginal cost of obtaining it. Nevertheless, once additional capacity investments have been carried out, and considering that indivisibilities will again lead to excess capacity, the efficient use of resources will indicate a need to charge lower fares. Hence, an efficient price system usually will lead to low revenue levels. This aspect of capacity is troublesome for the design of an RPI-X formula, because this regulatory system imposes rigidities that do not allow the necessary fluctuations to charge efficiently, nor do they permit the airport to break even.

Notably, when privatized airports forecast future investments, they take into account the actual price system upon which regulation is applied. Therefore, such a price system has to be consistent with the coverage of additional investment costs. Establishing a regulatory system that permits private airport operators to cover actual costs as well as those generated by future investments is needed. The British experience with regulation was such that the regulator was unable to design a regulatory mechanism that allowed investment decisions to rest entirely in the hands of private concessionaires. The regulator had to intervene to evaluate the impact of price regulation on investment plans. In this sense, the regulator adopted a managerial role.

Design of the Regulatory Mechanism

The British experience with regulation indicates that price cap regulation may impose certain risks on the regulated firm, making profits more volatile. This implies that regulated prices have to be frequently revised. Therefore, the regulator cannot establish a unique limit that is binding for a substantial period of time, and consequently, the main advantage of this regulatory procedure cannot be properly exploited. Apart from congestion and externality problems, complications also relate to the implicit incentive to degrade quality of service to increase profit margins. The lack of incentive to invest in new capacity further complicates regulatory prospects.

In the United Kingdom, the regulator has frequently had to intervene to compensate for the effects of the price cap formula. For example, adjustments in capital expenditures are often needed, and additional security costs need to be passed on to users. Other adjustments necessitated by inaccurate traffic forecasts that affect factor *X* are also common. If traffic increases are markedly above predicted levels, incrementally increasing investment expenditures may be necessary to avoid congestion problems. This would have clear repercussions on the airport's financial results. In other words, the regulator is often compelled to apply an ad hoc regulatory price mechanism.

An ad hoc regulatory mechanism may be a partial solution to the troubles that arise from purely applying a price cap. In this sense, Forsyth (1997) proposes using a mixed system, designed to combine regulation through the RPI-X formula with the rate of return. Fares are established with reference to the price cap formula and real airport costs. The weights given to each of these elements depend on the importance of different sources of inefficiency. For example, if quality is a serious problem, more emphasis is placed on airport costs. Airports are allowed to recover a great deal of the costs incurred by the provision of better quality of service. This mixed rule opens up the possibility of adjusting airport gains and losses in an ad hoc manner. Furthermore, it softens the crucial aspect of establishing an initial price upon which the regulatory mechanism is applied.

Hence, applying a mixed regulatory system in the airport industry may be desirable. This means a more active role for the regulator, however, because it is not possible to simply establish price regulation and leave the airport to make the rest of the decisions. The regulator needs to establish the necessary capacity at congested airports and, perhaps, the creation of a slot market. It should also estimate noise costs, establish charges for their internalization, and try to reconcile allowable noise levels with airport capacity. Finally, as a result of the importance of these quality-related aspects and the presence of these externalities at airports, the regulator has to directly decide on industry investment plans.

Quality and Safety Regulation

The main reason for regulating quality is market failure. Consumers are imperfectly informed about the quality of products at the time of purchase, and they are therefore unable to distinguish a poor-quality provider from a good-quality one. In general, regulation is needed to overcome this informational asymmetry. Nevertheless, the quality outcome may differ with the type of market and the temporal dimension. In competitive markets, firms that produce low-quality products and sell them at high-quality prices will acquire a bad reputation and will be excluded from the market (Klein and Leffler 1981). In monopoly situations, the quality of the product is always lower than in a perfect information setting. Imperfect information causes quality deterioration (Shapiro 1982). Regulators, however, face similar asymmetric information problems regarding product quality.

As discussed previously, privatized airports are usually subject to a regulatory pricing mechanism. Quality regulation is less common, despite the likelihood for exploitation of monopoly power in some airport operations.

For instance, the BAA is subject to price capping,[14] but it does not have to comply with a level of quality specified by the regulator. The BAA itself tracks its quality by periodically conducting quality survey monitoring. Being subject to CAA scrutiny seems a sufficient incentive for high standards of quality without any specific regulatory provision. Nevertheless, the BAA and the airlines agree on the level of service to be provided. The main areas of discussion usually are check-in, security queues, jetty availability, stand availability and cleaning, project development, and departure and baggage transfer. The final service agreement includes performance measures, service standards, and compensation in cases of nonfulfillment.

Monitoring Quality: The Case of BAA Self-Regulation

To evaluate quality performance at airports, distinguishing between the different recipients of airport services and the different ways of assessing quality is necessary. The main airport customers are the airlines, which in turn depend on paying passengers. For this reason, performance measure standards must distinguish between services directly provided to passengers and those intended for airlines. At the same time, two main approaches assess quality. The first is subjective, based on quality surveys that capture the quality perceptions of passengers and airlines. By contrast, more objective approaches measure performance in relation to standards (see table 3.11).

PASSENGER SERVICES. As mentioned above, the BAA controls the quality of passenger services through quality survey monitoring (the subjective approach). It measures passengers' perceptions of services in departures,

Table 3.11. *Elements for Quality Assessment at Airports*

	Recipients of airport services		
	Passengers	*Airlines*	*Others*
Alternatives for quality assessment	*Subjective approach*: Quality survey monitoring *Objective approach*: Establishment of standards and measurement of performance		

Source: Authors.

14. Price caps might induce quality cost cutting, as operators choose to reduce quality, and hence costs, rather than increase efficiency.

arrivals, and retail areas. It interviews more than 250,000 passengers each year. The interview takes 8 to 12 minutes, and passengers assess services on a five-point scale from "extremely poor" (1) to "excellent" (5). At Heathrow, Gatwick, and Stansted, the BAA has collected information over a six-year period on customers' perceptions of 12 basic aspects of departure services and 7 basic aspects of arrival services. It interviews departing passengers as they enter the gateroom, arriving passengers as they exit the terminal. In a similar way, it also collects perceptions about various aspects of service and value for money at retail outlets, car parks, and restaurants.

Tables 3.12 and 3.13 present results for those areas that are common to all three BAA airports. Each table shows the constituent factors for each airport, ranked according to the quality survey monitoring. Results show that, on average, passengers perceive most areas to be at least "average." Many areas are ranked between "good" and "excellent," and no areas are assessed as "extremely poor." Overall, Stansted scores consistently well, and Gatwick scores slightly better than Heathrow, although Heathrow has shown more improvement than Gatwick since 1991. In addition to recording subjective measures about passengers' perceptions, the BAA also has established various performance standards. Table 3.14 presents an example of check-in queue targets. Services provided to passengers are sometimes perceived as inadequate. Airports often devise a mechanism to treat complaints. The sensitivity of airport authorities to complaints, however, depends on their monopoly power and regulatory provisions.[15] Passenger complaints and suggestions may arrive in a variety of forms: comment cards, letters, phone calls, e-mails, or in-person visits. Their processing and treatment may be subject to regulation. Usually, a customer service department handles complaints, but the regulator may be the ultimate arbitrator. Additionally, targets may be fixed for prompt responses.

AIRLINE SERVICES. To completely assess quality, one must also take into account services directly provided to airlines. Although the BAA is not subject to quality standards, some airlines have requested the MMC to establish standards regarding the availability of key operational equipment such as baggage belts, jetties, stands, moving walkways, and lifts.

Despite a lack of quality regulation, the BAA makes direct measurements of its service delivery by recording objective data on the availability of critical equipment. Table 3.15 shows 24-hour availability data for passenger-sensitive equipment from April 1995 to March 1996. Other performance

15. Monopoly power here refers to the existence of competing airports.

Table 3.12. Quality Survey Monitoring Scores: Departure and Arrival areas at Selected BAA Airports, 1995/96

Items studied	Heathrow	Gatwick	Stansted
Departures			
Security queue	4.1	4.2	4.4
Telephones	4.0	4.0	4.1
Check-in queue	4.0	4.0	4.3
Departure lounge cleanliness	4.0	4.1	4.5
Flight information	3.9	4.0	4.0
Toilets	3.9	4.0	4.4
Trolleys	3.9	3.9	4.2
Airside seating	3.7	3.9	4.2
Announcements	3.7	3.7	4.0
Check-in crowding	3.6	3.8	4.1
Landside seating	3.5	3.8	4.1
Departure lounge crowding	3.5	3.8	4.3
Average	3.82	3.93	4.22
Arrivals			
Immigration queue	4.2	4.3	4.5
Disembarkation	4.0	4.0	4.1
Trolleys	3.9	3.8	4.2
Telephones	3.9	4.0	4.2
Baggage claim queue	3.8	3.9	4.0
Toilets	3.8	3.9	4.4
Concourse crowding	3.5	3.8	4.3
Average	3.87	3.96	4.24

Note: A score of 1 is "extremely poor," 2 is "poor," 3 is "average," 4 is "good," and 5 is "excellent."
Source: Monopolies and Mergers Commission (1996).

indicators the BAA developed are the number of faults per unit (a measure of the effectiveness of preventive maintenance) and the time to site and time to repair (measures of reactive maintenance). The set target is repairing 95 percent of faults within four hours. Table 3.16 shows average fault repair times for passenger-sensitive equipment. Other aspects considered are the percentage of passengers boarding or disembarking via jetty, coach, or steps (table 3.17); planned and unplanned stand outage (in hours per month); and maximum delivery times for baggage (table 3.18).

A regulator that is concerned about the exploitation of monopoly power must also consider a mechanism for registering airline complaints in the

Table 3.13. Quality Survey Monitoring Scores: Retail Value for Money at Selected BAA Airports, 1995/96

Items studied	*Heathrow*	*Gatwick*	*Stansted*
Duty-free shopping	3.8	4.1	4.1
Tax-free shopping	3.7	3.8	3.8
Other shopping	3.6	3.7	3.7
Catering	3.4	3.5	3.5
Bureau of change	3.3	3.5	3.4
Long-term parking	3.3	3.6	3.4
Short-term parking	2.7	3.1	3.5
Average	3.4	3.6	3.6

Note: A score of 1 is "extremely poor," 2 is "poor," 3 is "average," 4 is "good," and 5 is "excellent."
Source: Monopolies and Mergers Commission (1996).

Table 3.14. BAA Check-In Queue Targets

			Length (persons queuing)		
			Gatwick		
Maximum wait time (minutes)			*Scheduled*		
Heathrow	*Gatwick*	*Stansted*	*Short-haul*	*Long-haul*	*Charter*
20	20	15	10	18	18

Source: MMC (1996).

event that the airlines are disappointed with airport services. For instance, in the case of the BAA, the CAA is responsible for addressing the complaints of airlines and other agents such as tour operators and concessionaires. Other airports that might feel damaged by anticompetitive practices may also refer to these authorities or to the MMC.

Many of the most crucial aspects of airport operations are not always the direct responsibility of the airport authority. The punctuality of aircraft landings and takeoffs is also determined by visual and approach air traffic services. To keep up with published timetables, the airport authority and the air traffic control must closely coordinate, particularly when they belong to different organizational bodies.

All the above-mentioned variables represent possible regulatory quality targets (even in terms of scores or standards). Standards may be applicable when full divestiture has been applied or when a concession contract

Table 3.15. *Percentage Availability of Critical Equipment at Selected BAA Airports, April 1995 to March 1996*

Airports	*Departure baggage systems*	*Passenger lifts*	*Loading bridges*	*Passenger conveyors*	*Escalators*
Heathrow					
Terminal 1	97.8	99.2	99.0	99.1	98.9
Terminal 2	98.8	99.5	99.3	—	99.3
Terminal 3	98.4	99.4	98.7	98.5	99.3
Terminal 4	98.6	99.5	99.4	99.7	99.8
Gatwick					
North Terminal	98.1	99.4	98.5	99.2	99.5
South Terminal	97.5	99.2	97.9	98.8	99.2
Stansted	99.4	99.4	99.5	—	99.9

— Not available.
Source: Monopolies and Mergers Commission (1996).

Table 3.16. *Average Fault Repair Times for Critical Equipment at Selected BAA Airports, April 1995 to March 1996*
(hours)

Airports	*Departure baggage systems*	*Passenger lifts*	*Loading bridges*	*Passenger conveyors*	*Escalators*
Heathrow					
Terminal 1	1.52	3.18	2.31	1.92	1.79
Terminal 2	0.14	2.35	1.31	—	3.24
Terminal 3	0.55	4.53	4.24	4.45	1.82
Terminal 4	0.63	3.75	0.83	0.68	1.09
Gatwick					
North Terminal	0.92	2.04	1.25	1.49	1.29
South Terminal	1.60	1.97	6.27	1.57	1.81
Stansted	0.12	1.64	0.46	—	0.31

— Not available.
Source: Monopolies and Mergers Commission (1996).

Table 3.17. Average Levels of Pier Service at Selected BAA Airports, 1995/96

	Percentage of passengers boarding/disembarking via:			
Airports	*Arrival/ departure*	*Jetty*	*Coach*	*Steps*
Heathrow				
Terminal 1 Domestic[a]	Arrival	87	9	4
	Departure	89	7	4
Terminal 1 International	Arrival	79	17	4
	Departure	81	15	4
Terminal 2	Arrival	95	2	3
	Departure	94	2	4
Terminal 3	Arrival	89	7	4
	Departure	90	4	6
Terminal 4	Arrival	94	3	3
	Departure	93	4	3
Gatwick				
North Terminal	Arrival	83	13	4
	Departure	75	21	4
South Terminal	Arrival	83	3	14
	Departure	83	3	14

a. Including Channel Islands and Ireland.
Source: Monopolies and Mergers Commission (1996).

Table 3.18. Standards for Maximum Baggage Delivery Times at Selected BAA Airports
(minutes)

Airports	*First bag*	*Last bag*
Heathrow		
Terminal 1	16–20	30–34
Terminal 2	21	25
Terminal 3	24–28	49-53
Terminal 4	11–20	22–41
Gatwick		
North Terminal	20	35
South Terminal	20	35
Stansted	15	33

Source: Monopolies and Mergers Commission (1996).

is intended. The regulator should study the convenience of intervening to fix quality levels. A scrutiny mechanism and agreements with air transport carriers about prices and corresponding quality levels may be adequate to ensure high standards of quality.

Safety and Externalities

Airport safety plays an important role in determining quality. Its objective is to ensure that passengers have a normal wait and flight, with a minimal possibility of a terrorist or criminal attack. These safety standards and procedures impose costs on airlines and passengers. Table 3.19 shows different components of the airport security system.

Security queues are considered an important determinant of airport quality. The BAA reports that among Heathrow, Gatwick, and Stansted airports, 95 percent of passengers waited fewer than 10 minutes and 90 percent waited fewer than 5. Airlines suggest that a maximum waiting period of 5 minutes for a security search at London airports is desirable.

In economic jargon, externalities are considered a market failure, hence intervention is regarded as necessary wherever they appear. The main negative externalities at airports are noise, congestion, and pollution, as discussed earlier. Traditionally, airport operators and the corresponding regulators have left externalities aside. It is only recently that have they started worrying about their environmental impact. Today, it is common for airports to ban night operation or to restrict it to less noisy aircraft. Charging noisier planes higher fares is another technique to reduce the social cost of noise. Peak pricing is also spreading as a practice for relieving congestion. Air pollution has not gotten much importance.

Increasing sensitivity about environmental concerns has led to special treatment for externalities in most infrastructure project contracts. Usually, an environmental impact study is required as a prerequisite for airport infrastructure construction. This study should also consider monitoring possible negative impacts during the operation phase. In general, the environmental impact study will reflect environmental law.

Investment Obligations

As shown earlier, the possibilities are numerous for private sector participation in airports. If this involvement does not fulfill its long-term objectives to maintain the facilities and invest in the future, however, airports

Table 3.19. Components of the Airport Security System

Security component	*Functions*
Predeparture gate screening	Screening passengers, body search, screening airport and airline personnel, X-ray inspection of carry-on luggage.
Parked aircraft control	Screening airport and airline personnel, alarm systems for parked aircraft, aircraft security survey.
Aircraft movement	Screening airport and airline personnel, alarm systems for parked aircraft, aircraft security survey.
Crew screening	Background checks, training, predeparture screening.
Ramp security	Surveillance of jetway access, ramp doors, alarm systems, fire sensors and protection, screening personnel.
Perimeter security	Fencing, posts, gates and other openings, light placement and protection.
Terminal security	Surveillance of jetway access, ramp doors, alarm systems, fire sensors and protection, screening personnel.
Passenger screening	Visual, body searching, X-ray inspection, location.
Passenger flow control	Flow holding, camera surveillance, predeparture screening.
Baggage and cargo screening	X-ray inspection, carry-on luggage screening, luggage surveillance from drop-off to loading, personnel screening.
Intelligence and communications	Telephone and radio communications, emergency power, bomb threat contingency plans, evacuation plans.

Source: Fleming and Ghobrial (1993).

could deteriorate or become obsolete. Fortunately, this has not been the case in airport infrastructure concession contracts.

On the contrary, investment plans are usually an essential part of the contract. For instance, a recently prepared concession contract for the operation of Argentine airports required the operator to present a detailed investment plan. The concessionaire is obliged to invest a minimum of more than US$2 billion in addition to other planned investments (including a new airport for Buenos Aires). Such a plan must clearly specify in physical and monetary terms the works that will be carried out during the concession period.

BAA investment plans are also subject to CAA scrutiny. The BAA periodically presents projected investments, which are expected to be broadly in line with reality. Additionally, the BAA is required to consult airlines on future development plans.

Performance Indicators

The privatization of a firm leads to an increase in productive efficiency, because in the absence of regulation, the firm pursues profit maximization. If the firm also exerts monopoly power, however, it is possible that allocative efficiency is reduced as well. In this case, regulation could be a means for limiting market power, although it may also affect economic efficiency. Its impact would depend on the implemented regulatory system. In the airport industry, most regulation takes the form of price intervention. If charges are established independent of profits, productive efficiency is feasible, although prices are usually fixed in such a way that firm profits are under control. Price controls permit an improvement in allocative efficiency by reducing monopoly power. They can also reduce productive efficiency, however (Forsyth 1997). Hence, regulation may affect the economic efficiency of the airport industry, making it necessary to develop performance indicators that monitor airport activities liable to be affected by regulation.

Evaluating airport efficiency is not a trouble-free task. The geographic, economic, political, and social features of the airport's region complicate any assessment of industry efficiency. Doganis (1992) points out that evaluations tend to be based on profit margin analysis. Obviously, this criterion is inadequate because it does not incorporate any information about the resources that go into obtaining such a margin. Therefore, establishing indicators is essential both to assess the effectiveness of resource utilization and to serve as control tools for airport managers seeking to identify problematic areas that require prompt corrective measures. Indicators are also a great help for governments concerned with regulation. For example, they could be used to ensure that national resources are being used in the most efficient way, that airports are not exerting their monopoly power, and that they are providing the required services at reasonable prices.

Given the trend toward airport privatization, government responsibility should be directed at the creation of a regulatory policy that channels private sector performance to match public interests. In this sense, using indicators may contribute to evaluating this accomplishment. In the British case (BAA), privatization brought clear management efficiency improvements, mainly at airports in London. Nevertheless, the MMC may carry out controls at these airports to determine if their monopoly power is being exerted against public interests. The main criticisms relate to the following three areas: (a) service quality, (b) fare levels and structure, and (c) investment levels and quality. The commission also controls other elements that are not subject to regulation, such as rents, licenses, and commercial concessions. This was of great importance due to the tough regulation applied on BAA airports, which resulted in aeronautical charges that were below associated costs and a need to cross-subsidize these services with revenues arising from commercial activities. As a consequence, users had to pay monopoly prices in commercial areas to complement the aeronautical-side deficit, and thus subsidize air transport carriers.

According to the literature on airport industry management, financial and economic indicators are usually the most utilized (Ashford and Moore 1992; Doganis 1992). Given that one of the main objectives of a private firm is cost minimization, a useful measure of efficiency must cover financial aspects. Economic objectives such as input productivity are also important to any industry. Therefore, a menu of economic indicators is also necessary. Nevertheless, as indicated above, these indicators should be complemented by other measures that allow the evaluation of airport services and activities that may cause problems for users. Elements such as quality of service and negative airport externalities should be considered as well. For instance, waiting times and congestion in the terminal building are of primary importance to users' perceptions of quality of service.

Elements That Determine Indicator Design

Before proposing a set of indicators, we should note that some aspects directly affect their utilization. First, airports develop similar activities for different objectives. In addition, these objectives may conflict with one another. For example, an increase in airport runway capacity through the establishment of additional approach routes also raises the level of noise. Furthermore, each airport has a different social, economic, and political environment. For this reason, proposing a set of indicators without taking the special features of each airport into account is a risky task. Indicators ought to adapt to the social, economic, and political characteristics of each airport. Disparities among airports need to be considered when fixing reference standards.

Second, the information used in calculating the indicators must comply with certain requirements, such as easy access, clarity, and accuracy, so that nonspecialists can understand them. This should cover most aspects of airports (ICAO 1991). Such an evaluation and control process should be carried out as an integral part of the airport planning program, not as an assessment of the private manager's responsibilities. A conflict due to information asymmetry exists, however, because the private operator has an incentive to hide relevant information from the regulator. This situation might be softened by reasonable service standards and periodic controls that allow continuous supervision.

A troublesome element in the evaluation of airport performance and productivity is defining the output used. An airport output is not homogenous. It can be defined in terms of the number of planes, passengers, and cargo volumes. Each of these output measures, however, only relates to a part of the infrastructure. Runways relate to the number of landed aircraft, and terminal building size depends upon the number of passengers and cargo processed. Therefore, no single measure can entirely explain airport costs and revenues.

Doganis (1992) argues that the choice of output must be in accordance with its economic importance in terms of revenues and cost generation. In this sense, for most airports around the world, the greatest proportion occurs in activities developed in the terminal building, such as passengers and cargo handling. Therefore, an output measure that combines both variables would cover the largest proportion of airport revenues and costs. Passengers and cargo volumes are an indirect measure of the total number of processed aircraft. Actually, one frequently uses the variable work-load units (WLU) as an adequate measure of airport output. A WLU corresponds either to a passenger (80 kilograms average weight plus 20 kilograms of luggage) or to 100 kilograms of cargo. Notably, though, one passenger and one unit of cargo do not require the same use of physical and financial resources, nor do they generate the same revenue. By contrast, some indicators demand a given output measure. For example, when assessing revenues that arise from commercial activities, using traffic units in the denominator does not make sense.

Obtaining output measures is relatively easy; consequently, obtaining the necessary data required by the indicators should not be any problem. Input measures, however, cause more serious problems. The most important inputs at airports are labor and capital. The easiest measure of the former is the number of workers. This is not homogenous, however, because it includes both part-time and full-time personnel, whether skilled (technicians and managers) or unskilled. Because different types of workers carry out different tasks at airports, developing a more comprehensive and accurate measure for determining the labor input is necessary. A solution may be to consider the financial value of the input (Doganis 1992). Nevertheless, this measure also presents considerable problems because it reflects not only the quantity of the input applied, but also the relative wage differentials among airports. This further complicates the use of indicators that serve as standard references. Consequently, using the number of workers as a measure of labor input is advisable. The number of workers, however, must be properly classified in order to evaluate a particular area. For example, only those workers directly involved in aeronautical activities should be used in the denominator when calculating revenue per unit of labor input.

Regarding the capital factor, the situation is even more complicated. This is essentially due to the diverse nature of capital inputs. For instance, the differences between small capital resources with short economic lives and large, long-term investments (runways and buildings) make posterior input allocation very difficult to measure. The ICAO recommends using asset values to measure capital. The existence of diverse accounting methods, however, means one must be careful. For example, if capital goods investments are financed by government funds, depreciation is likely not entered into the account. This procedure is common at airports that have traditionally operated as public firms. Determining asset value at these airports is misleading because of the variation in accounting practices. Nevertheless, no alternative exists to trusting financial measurements of capital. For a more reliable evaluation of inputs, the whole industry needs to adopt a common accounting system.

Infrastructure Performance Indicators

This work presents performance indicators commonly used in the airport industry. In some cases, however, it might be necessary to make a selection or an ad hoc design according to the special airport features and services that need to be assessed. Although the proposed list is not exhaustive, it covers those aspects or areas that might be problematic for regulators and managers. Particular areas are more likely to infringe upon public interest. For example, at airports subject to price regulation, problems arise regarding incentives to invest in new capacity and with the quality of service. This is the result of strong operator tendencies toward reducing costs at the expense of service quality. Therefore, having a set of financial and economic indicators available that helps analyze airport performance is important. These could include cost coverage, profitability, asset investments, and the use of available resources.

Table 3.20 presents a set of financial indicators. The first group, the *strategic indicators,* is needed to evaluate the medium-and long-term effects of policies, such as return on capital investments. Second, *other financial indicators* include measures such as cash flow, and they accurately evaluate the day-to-day financial situation of the airport.

Table 3.21 shows economic efficiency indicators. These are classified into six distinct categories: overall cost performance, labor productivity, productivity of capital employed, revenue-generating performance, performance of commercial activities, and overall profitability. To assess the economic efficiency of an airport through time, or to check whether regulated standards are being met, specific indicators are required. For example, we may need to explore labor and capital productivity to determine the most efficient use of resources. Alternatively, to gauge the performance of commercial areas, having specific revenue indicators is necessary (Doganis 1992).

Revenues from leasing, licenses, and concessions derive from activities that are not subject to regulation. These activities, however, must also be evaluated. For example, if rents paid by commercial area tenants are excessive in comparison to other rents in the market, imposing controls may be necessary. Aeronautical charges are determined by the single till approach, in which airport costs and revenues are viewed by taking all

Table 3.20. *Financial Performance Indicators*

Type	*Examples*
Strategic indicators	Return on capital investment
	Payback period
	Current assets/liabilities
	Self-financing ratio
	Ratio of debtors to creditors
Other financial indicators	Cash flows
	Revenue flows
	Expenditure flows
	Actual and budgeted revenues and expenditures
	Outstanding debtors and location of debt

Source: Lemaitre (1997).

Table 3.21. Economic and Productivity Indicators

Type	*Examples*
Overall cost performance indicators	Total costs per WLU (after depreciation and interest) Operating costs per WLU (excluding depreciation and interest) Capital costs per WLU Labor costs per WLU Labor costs as percentage of total costs Capital costs as percentage of total costs Aeronautical costs per WLU Capital costs to value added ratio Labor costs per employee
Labor productivity indicators	WLU per employee Total revenue per employee Value added per employee Value added per unit of staff plus capital costs Value added per unit of staff costs
Productivity of capital employed	Value added per unit of capital costs WLU per £1,000 net asset value Total revenue per £1,000 net asset value
Revenue generation performance	Total revenue per WLU Adjusted revenue per WLU Aeronautical (or nonaeronautical) revenue as a percentage of total revenue Aeronautical revenue per WLU Nonaeronautical revenue per WLU
Performance of commercial activities	Concession plus rental income per passenger Concession revenue per passenger Rent or lease income per passenger Concession revenue per m^2 Rent or lease income per m^2 Airport concession revenue as percentage of concessionaires' turnover
Profitability measures	Surplus or deficit per WLU Revenue to expenditure ratio

Source: Doganis (1992).

services into account. Aeronautical charges are fixed to permit a given profitability level that, in turn, depends on previous cost and revenue estimations. Once the regulatory pricing formula is in place, the private operator could increase rents above those charged in commercial areas, therefore acting against the public interest.

As already mentioned, if airports are subject to price regulation, they may also be tempted to reduce service quality in order to reduce their costs. Therefore, investigating users' perceptions about the services provided at airports is crucial. Before carrying out a quality assessment, defining a standard level of service that is both feasible and reasonable is necessary. These standards allow the airport regulator, under the threat of penalty, to demand the attainment of a certain level of service.

In the British case, air carriers have argued the need to reach an agreement regarding the standard level of services, as well as for provisions that entitle them to compensation in the event of nonfulfillment. They maintain that any deviation from standards affects their service quality and that without compensation, this mechanism is not effective. The BAA, however, argues that airport services are jointly provided by airport operators and airlines and, therefore, the level of service does not depend entirely upon its performance, but also upon air carriers and handling staff. Carriers, in turn, argue that penalties must be applicable only to the BAA, because airlines operate in a competitive environment that gives them strong incentives to maintain and improve their quality. Either way, a key aspect of this compensatory mechanism is identifying who is responsible for not achieving the standards. Reaching an agreement regarding quality standards is essential to guarantee a certain level of service within the context of regulation.

The procedures are complex for evaluating the factors that determine the service levels in terminal buildings. This leads to the use of variables such as time of service and level of congestion as proxies for the quality of services provided. Table 3.22 shows a set of quality indicators for most of the conflictive aspects of airport activities.

A study carried out at the Birmingham airport (see Mumayiz and Ashford 1986) established that users' perceptions of time of service depended on the type of market. For European flights, a check-in waiting time of 7.5 minutes or less was considered satisfactory, and a time equal to or greater than 14 minutes was perceived to be intolerable. For charter flights, these limits were 11 and 21 minutes, respectively. According to the same study, a general waiting time of not more than 12 minutes indicates a satisfactory level of service.

A trade-off can be found, however, between the level of service offered and its costs. The higher the level of service, the higher the amount of resources required. If we could identify all or some costs associated with the time wasted

Table 3.22. Quality of Service Indicators

Type	*Examples*
Delays	Time of service: check-in time, luggage delivery time, and so forth Waiting time Waiting time variability
Service reliability	Baggage service reliability Number of luggage incidents Number of passengers delayed at departures Required time before departure Connecting time
Costs	Costs for passengers of food and drink Departure fee Connecting fee Other services fees
Comfort and entertainment	Crowding at the terminal: number of square meters per occupant Clarity and level of noise Temperature and humidity levels Choice of leisure activities Sociability Cleanliness Air pollution

Source: Adapted from Lemaitre (1997).

by passengers in queues, and the economic resources wasted as a consequence of this wait, assessing the losses arising from the level of service provided would be possible. In summary, establishing an inadequate level of service could negatively influence users and even airports. An example is the check-in service: the more time passengers spend in front of check-in counters, the less time available for shopping in the commercial areas.

Using indicators as tools for assessing a given activity is ineffective if no reference standards delineate acceptable performance margins. Once taking the particular features of each airport into consideration, however, these desirable or best practice reference standards should only be regarded as provisional guides. No unique optimum level exists for a given indicator. The appropriate and optimal reference level depends on the circumstances of each airport. Furthermore, conflicts could be possible among the different objectives pursued. For example, an improvement in the level of quality may require a substantial increase in costs, which would eventually be translated into higher fares. With these arguments considered, reconciling the establishment and implementation of indicator reference standards is important. Table 3.23 gives some examples of indicators and their associated standards.

Doganis and Graham (1995) have carried out evaluations of the economic and commercial aspects of 25 European airports through the application of a set of performance indicators.[16] The authors emphasize the comparison problems due to differences in the activities developed at each airport in the study, which they tried to lessen through corrections that consider the whole group as operators of the same activities.[17] The sample includes airports with different ownership structures and of varying sizes: private airports such as Glasgow, partially privatized ones such as Copenhagen, publicly owned but commercially oriented airports such as Geneva, and airports like Stockholm, which is part of the Swedish Civil Aviation Authority. The main objective of this study was to analyze the trends and development of industry performance and identify the relationship between profitability and type of airport. Table 3.24 summarizes the results.

Finally, note that the use of indicators is complicated because of differences in the types of services developed at airports, in the degree of public intervention, in accounting systems, in financial sources, in subsidies, and in standards. All of these elements, combined with the geographic, economic,

Table 3.23. *Examples of Reference Standard Levels*

Type	*Indicator (example)*	*Best practice*
Financial	Return on capital investment	>1.0
Labor productivity	Passengers per employee	2,000 to 5,000
Service quality	Number of square meters per occupant at peak hours	25–35 (international); 16–20 (national)

Source: Authors.

16. Airports included in the study were Amsterdam, Barcelona, Basel, Mulhouse, Bilbao, Birmingham, Cardiff, Copenhagen, Dublin, Düsseldorf, East Midlands, Frankfurt, Gatwick, Geneva, Glasgow, Heathrow, Lisbon, Madrid, Manchester, Milan, Newcastle, Nice, Oslo, Stockholm, Vienna, and Vigo.

17. Adjustments that the study carried out indicate that results must be carefully analyzed. Comparability problems are still present, and consequently, one should consider each airport in its own context.

Table 3.24. *European Airports: Best and Worst Practice Values*

Indicators	*Worst practice*	*Value (US$)*	*Best practice*	*Value (US$)*
Cost indicators				
Total costs per WLU	Basel-Mulhouse	14.3	Oslo	2.94
Operating cost per WLU	Vienna	10.58	Oslo	1.94
Capital costs per WLU	Basel-Mulhouse	6.51	Oslo	0.99
Labor costs per WLU	Vigo	7.07	Oslo	0.73
Productivity indicators				
WLU per employee	Vigo	4,367	Oslo	48,808
Total revenue per employee	Vigo	17,930	Oslo	389,053
Value added per employee	Vigo	9,280	Oslo	329,997
Value added per unit of staff costs	Vigo	0.30	Oslo	9.23
Revenues indicators				
Total revenues per WLU	Vigo	4.11	Vienna	19
Aeronautical revenues per WLU	Vigo	2.38	Vienna	9.9
Nonaeronautical revenues per WLU	Lisbon	1.67	Vienna	9.1
Rent and lease income per passenger	Vigo	1.55	Gatwick	8.8
Concession revenue per passenger	Vigo	1.12	Gatwick	7.65
Financial indicators				
Revenue to expenditure ratio	Vigo	31	Oslo	272

Source: Doganis and Graham (1995).

social, and political characteristics of the airport region, hinder the assessment of airport performance.

Conclusions

Public utilities infrastructures have been traditionally regarded as natural monopolies. Public ownership was unquestionably the management style in terms of fares charged and quality provided. The underlying assumption was that both variables were set so as to maximize social welfare. Airport infrastructures provide an example of what improvements can be made when the economic system no longer can bear the burden of inefficiency. Taking into account the variety of activities carried out at an airport, the range of possibilities for private sector involvement may be as wide as the range of airport activities themselves. Nevertheless, no best practice model has emerged, and each country implementing airport privatization measures has selected the scheme that most adequately suits its needs. The alternative of concessions (in any of its variants), however, appears to be the emerging model. It allows the government to retain the property and facilities at the end of the concession period, and furthermore, it provides a financial windfall for governments with restricted budgets. Still, airport privatization processes are not cost-free. If public monopolies are being turned into private monopolies, a regulatory cost will be charged. Either discretionary or contractual, regulations will mainly affect airport charges, quality of service, investment obligations, externalities, and safety control.

4

Seaports

Lourdes Trujillo and Gustavo Nombela

The relevance of seaports to the efficient working of an economy cannot be understated, because all goods and passengers transported by sea require the use of ports' facilities. In most countries, maritime transport basically handles the export and import trade and, in some cases, also a large share of domestic trade. For long-haul shipments, with the exception of high-value and small-volume cargo, for which air transport offers the advantage of speed, there is no alternative mode of transport to ships.

The United Nations Conference on Trade and Development defines the role of a modern seaport as follows:

> Seaports are interfaces between several modes of transport, and thus they are centers for combined transport. Furthermore, they are multi-functional markets and industrial areas where goods are not only in transit, but they are also sorted, manufactured, and distributed. As a matter of fact, seaports are multi-dimensional systems, which must be integrated within logistic chains to fulfill properly their functions. An efficient seaport requires, besides infrastructure, superstructure and equipment, adequate connections to other transport modes, a motivated management, and sufficiently qualified employees.

This definition stresses one of the main characteristics of seaports: a seaport is not merely an organization that provides a single service, but multiple activities. Studying all those tasks in detail is therefore interesting to evaluate the most efficient provision of these activities from an economic point of view. Moreover, because all port activities take place in a limited area, studying how they are coordinated is also relevant, as is the role that port authorities—or any other responsible institutions—must play in regulating seaports' infrastructure and activities.

In recent decades we have witnessed profound changes in maritime transport that have modified the balance between capital and labor at seaports. Ports are now increasingly becoming capital-intensive industries, while in the past they used to be labor-intensive. This change has generated an excess of employees in most ports worldwide. The development of containerized transport is another factor that has significantly modified ports' operations. Containers have achieved significant cost reductions in cargo handling, but they have also imposed new needs on ports in terms of equipment (gantry cranes, improved pavements, and so on). However, the transport of large quantities of containers and bulk cargo has yielded economies of scale, which have led to the building of increasingly larger specialized ships that require new infrastructures and equipment.

All these technical changes have generated a highly competitive environment in the seaport industry, especially between large ports with facilities for serving regular deep-sea traffic from liners. Modern ports no longer have a monopoly over the transport of goods to neighboring regions (hinterlands). The development of integrated transport chains has reduced transport costs to such an extent that it is now often preferable for a shipper to use a distant port instead of a closer one, provided that the former has better facilities and connections than the latter. Therefore, modern ports must be extremely competitive to be able to offer optimal combinations of time and price for firms demanding their services.

Technological changes and the competitive environment have induced a reconsideration of the role that the public sector must play in the running of seaports. In most countries, public institutions have traditionally owned and managed ports. Public ownership in the industry was justified by the argument that seaports play a key role in national economies, and they exhibit characteristics that can easily provide firms with market power (such as specialized assets, sunk costs, indivisibilities, and economies of scale).

Even though the public sector has usually been present as port organizer, however, it is not evident that public organization of this industry is necessarily the best option. In particular, tighter public budgets and increasing fiscal needs have led many countries to seek private participation in seaports. Private involvement in ports is not new for the provision of services, because many firms were already present in ports around the world, but it is innovative in the construction of port infrastructure.

International experiences have shown that private participation in both these aspects (operations and infrastructure) has significantly improved the outcomes of some seaports. These experiences make a case for a revision of the traditional organization of seaports worldwide, changes that will prepare ports for a more competitive market and less financial help from governments

This chapter offers a revision of characteristics of all different services that seaports provide, and it describes the approaches used worldwide to introduce private participation into the port industry. The challenge that modern ports now face is to design more adequate regulatory mechanisms to guarantee efficient outcomes in a context of tight public budgets, particularly in developing countries. No universal answers apply to every port; therefore, this chapter provides a panoramic view of the feasible models a port can follow, and the best practices observed worldwide.

Characteristics of Seaports' Services: The Multiproduct Nature of the Activity

In broad terms, a seaport can be considered a single organizational unit that provides a service to ships. When its internal workings are analyzed in detail, however, multiple services clearly are being produced and demanded within a port area (services to ships, to cargo, and to passengers). Even for such a service as cargo handling, technologies can vary enormously depending on the type of cargo, up to the point that, for example, container loading can be regarded as a different service from bulk cargo handling. Therefore, instead of a single unit, a seaport is better characterized in economic terms as a multiproduct organization.

Two basic characteristics define the organization of seaports' activities. The first is that the infrastructure in which these activities are performed—berths, quays, storage areas, and so forth—is expensive to build (see table 4.1) and it exhibits the problem of indivisibility: enlarging a port in a continuous way is not possible. Port infrastructures must be built with determined minimum dimensions, and in general, their full design is strongly conditioned by the physical characteristics of the coast where the port is located.

The second characteristic is that because of high construction costs and physical conditions, the areas available for performing seaport activities are generally limited. This space limitation implies that the number of operators that can provide services within a port area is, by definition, reduced. In particular, depending on the port's total size, some small ports

Table 4.1. Typical Civil Works Unit Costs for Port Infrastructure

Type	Cost
Dredging (confined space restricted by existent berth requirements)	$7.5/cu.m
Quayside (35m-wide berth)	$54,000/m
Container yard paving and infrastructure	$63.8/sq.m
Open storage yard	$55/sq.m
Sheds	$375/sq.m

Note: 1997 data, for a 14-meter draft, 500- to 600-m-long berth.
Source: Drewry Shipping Consultants (1998).

only have enough room for a single firm to provide services. Market size and physical restrictions are two factors that often preclude the possibility of competition at some ports.

Because many aspects are involved, dividing seaport activities among three groups is useful: (a) infrastructure; (b) services provided by the port, which require the use of the former; and (c) coordination among different activities performed at ports. The main characteristics of these three elements are analyzed below.

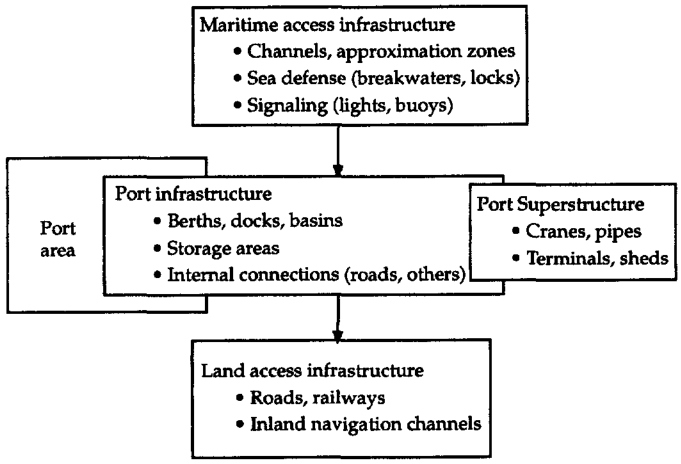
Seaport Infrastructures

The European Union uses an interesting definition of what is considered port infrastructure (European Parliament 1993). The port area is defined as a complex of berths, docks, and adjacent land where ships and cargo are served. To reach that area, infrastructures related to both maritime access (channels, locks, aids to navigation, and so forth) and land access (connections to roads, rail network, and inland navigation) are required. Figure 4.1 shows a scheme of the different types of required infrastructures at a port.

Therefore, the area where seaport activities take place encompasses both the infrastructure within the port (berths, quays, docks, storage yards, and so forth) and the so-called superstructure. Among the elements forming the superstructure, one can distinguish between fixed assets built on infrastructure (sheds, fuel tanks, office buildings, and so forth) and fixed and mobile equipment (cranes, van carriers, and other port equipment).

When discussing port infrastructure, precisely defining the included elements is convenient. As can be observed, infrastructures outside the port area are essential for the use of a port (maritime and land accesses). In general, port authorities have responsibility over some of the maritime access infrastructures—breakwaters, lights, buoys—and all elements within

Figure 4.1. Scheme of Seaport Structures



Source: Authors.

the port area. Meanwhile, the state or local government generally owns and maintains connections to land networks and the remaining forms of maritime access (channels, locks).

In almost all countries, port authorities have traditionally designed and maintained port infrastructure directly using public funds to finance the building of new infrastructure. People generally considered that the public sector should own these assets to avoid the risk of monopolization by private firms. A current global trend, however, is toward increasing participation of private capital in the construction of infrastructure, generally through the use of concessions.

Seaport Services

In addition to providing basic infrastructure for the transfer of goods and passengers between sea and land, different agents provide multiple services at ports. Some agents may even work outside the port area. These services cover all activities that connect port users to the port, from the moment that a ship approaches a port until it ends all its operations. During this period, agents provide services to the ship, to passengers, to the crew, and to cargo (De Rus, Román, and Trujillo 1994).

First, a group of services relates to berthing, which includes pilotage, towing, and tying. Port authorities can directly provide these services, or private firms can offer them. Pilotage is defined as those operations required for a ship to enter and exit a port safely, and it usually implies the presence in the vessel's bridge (or at least a radio contact) of an expert with sufficient knowledge of the zone to avoid risks. In some ports, pilots are independent private agents, licensed by the port authority, while in other cases they are public employees. Towage is the operation of moving a ship using small powerful boats (named tugs) to steer it more easily. Again, private firms could provide the services for these operations at some ports, while at other ports the port authority directly hires tugs and their operators.

One of the more important services provided to cargo ships is what is generically labeled cargo handling. This encompasses all activities related to the movement of cargo to and from ships and across port facilities. A historic separation exists between the movement of goods from the ship's side to safe storage within the vessel (stevedoring) and those movements from berth to ship's side (loading). This separation occurs as a result of different workers traditionally performing these operations. Today, however, specialized firms provide all cargo handling services, using equipment such as cranes and surface transport elements.

The cargo handling process varies according to the type of goods involved. The specialization of firms according to type of cargo is becoming a trend, because the equipment required can then be specifically designed to be highly cost-efficient. This leads to the formation of terminals, which are specialized berths in which operations concentrate on a given type of cargo. Container terminals constitute the best example of this trend, because the handling of containers requires large gantry cranes, and land storage is relatively easy with adequate trucks and lifts, but is highly space-consuming. All these factors make a specially designed berth more efficient for handling containers than general cargo berths.

Of the total costs involved in moving goods through a seaport, cargo handling charges are the most important (approximately 70 to 90 percent of total cost, depending on the type of goods). Therefore, this is one of the services that a regulator must closely supervise to ensure cost-efficient port operations.

Another type of service that port users demand is related to administrative paperwork and permits (sanitary certificates, import/export documents, taxes, and so on). Specialized agents or consignees usually perform this service, hired by shipping companies to prearrange paperwork and all matters related to the use of port facilities by a ship. Before a ship calls at a port, consignees work to ensure that all required services (handling, repairs, supplies, and so forth) for the ship are contracted to be performed in the shortest possible time.

Modern ports must have systems that minimize the paperwork burden for port users, because delays due to inefficient administrative procedures cause large economic losses for both shippers, which have to alter their productive plans when they do not receive their goods on expected dates, and shipping companies, which have to keep their ships in port longer than necessary. In the European Union, established guidelines promote ports' investments in developing electronic data interchange systems that are designed to speed up administrative paperwork and reduce wait times for ships and for the land transport that delivers goods to and from ports (European Commission 1997).

Finally, different agents and firms that work in the port area perform other ancillary services. These include all supplies for ships, of which fuel and water are probably the most important. Also included are services to crew members (such as medical) and general common services such as cleaning, refuse collection, and safety. Some ports also offer repair facilities to ships, which may involve the use of some special infrastructures.

In summary, a port offers many different services. A combination of public and private initiatives can perform these services, and several port models show how private participation is introduced. Table 4.2 shows a summary of all services described in this section. From the regulatory point of view, ensuring an adequate provision of infrastructure and cargo handling is the more relevant issue, because the efficiency of seaports hinges on these two services.

Coordination of Activities: Port Authorities

Many different activities are performed simultaneously within the limited spaces of port areas, as ships constantly enter, are serviced, and exit. Therefore, an agent is needed to act as coordinator to ensure the proper use of common facilities and to take care of safety and the general design of port facilities. In most seaports, an organization called the port authority performs this function. Generally, these are public institutions that represent local interests, but this configuration is not unique, and finding examples of purely private port authorities is possible.

One can characterize several types of port organization, depending on the role of port authorities. These are usually referred to as landlord port, tool port, and services port (Juhel 1997) as follows:

Table 4.2. Port Services

1. Infrastructure provision	
2. Berthing services	Pilotage Towing Tying
3. Cargo handling	Stevedoring Terminals Storage Freezing (fish, others)
4. Consignees	Administrative paperwork for ships and cargo Permits (sanitary, customs, and so forth) Service hiring
5. Ancillary services	Supplies Repairs Cleaning and refuse collection Safety

Source: Authors.

- *Landlord port.* In this model, the port authority owns port infrastructure and is also in charge of its management. Private firms that own the assets of the port superstructure and all equipment required for service provision (such as cranes, vans, and forklifts) provide the remaining services. Examples of this type of port organization are Buenos Aires (Argentina) and Rotterdam (Netherlands). This is generally the most common form of organization for large ports.
- *Tool port.* As in the landlord model, port authorities are the owners of infrastructure, but they also own the superstructure (such as buildings) and the equipment (such as cranes). Private firms provide services by renting port assets through concessions or licenses. Examples of this category are Antwerp (Belgium) and Seattle (United States).
- *Services port.* In this model, port authorities are responsible for the port as a whole. They own the infrastructures and superstructures, and they also hire employees to provide services directly. The port of Singapore used to be the best example of this type of organization, with its port authority (PSA) being the owner of all assets and providing all services. In 1996, however, PSA was split into two separate entities: Maritime and Port Authority, landlord and regulator, and PSA

> Corporation, the port operator. Even though the law now authorizes the entry of private operators and the privatization of PSA Corp., the port of Singapore is still run by public institutions. Its model of organization, however, is converging toward a landlord port model.

To establish a connection between the type of port and ownership, in landlord and tool ports, the port authorities generally are public institutions, while the port operators are private firms. Therefore, these two types could be classified as mixed ownership, because the basic infrastructure is generally public, but operators can own many elements of the port. Services ports are more likely to be privately owned, with a single private firm operating the port as a single unit.

Even though this is the general pattern, finding examples in the port industry in which ownership and mode of organization do not follow the above scheme is possible. For example, services ports that are fully public can be found—such as Singapore, although as mentioned above, this port will likely transform soon—as well as landlord ports in which infrastructure is privately owned. Table 4.3 shows the type of ownership for the 50 largest world ports according to container traffic. In this ranking, one can observe a trend in the seaport industry toward ports with mixed ownership, at least for large ports.

Some port facilities have traditionally been regarded as public goods (lights, access channels, and so forth). A seaport considered as a whole, however, does not exhibit public-good characteristics. For ports, excluding users is possible, and producing services to more users without increasing costs is not feasible. Therefore, seaports are organizations that from an economic point of view do not necessarily have to be in the public sector. They can be run as commercial institutions.

Some countries regard many port activities as public services in the sense that authorities consider that these services should be available to any user, but there is no reason for the public sector to be obliged to provide them directly. Only in particular situations, as in the case of extremely small seaports serving isolated communities, can conditions be found in which public intervention would be required to guarantee the provision of port services, because the seaport would be vital for the community's basic welfare. However, even in this case, a public company should not necessarily be strictly providing port services, but instead subsidized private firms could offer them through competitive bidding for concessions with negative payments.

Table 4.3. *Type of Ownership of 50 Main World Ports, 1997*

Port		TEUs thousand	Public	Mixed	Private
1	Hong Kong (China)	14,539		•	
2	Singapore	14,135	•		
3	Kaohsiung	5,693		•	
4	Rotterdam	5,445		•	
5	Pusan	5,234		•	
6	Long Beach	3,505		•	
7	Hamburg	3,370		•	
8	Los Angeles	3,000		•	
9	Antwerp	2,969		•	
10	Dubai	2,600		•	
11	Shangai	2,527	•		
12	New York/New Jersey	2,457		•	
13	Tokyo	2,383		•	
14	Yokohama	2,328		•	
15	Felixstowe	2,251			•
16	Keelung	1,981		•	
17	Kobe	1,944		•	
18	San Juan	1,781		•	
19	Bremen	1,526		•	
20	Colombo	1,687	•		
21	Kelang	1,684			•
22	T. Priok	1,671		•	
23	Algeciras	1,538		•	
24	Oakland	1,531		•	
25	Nagoya	1,498		•	
26	Seattle	1,476		•	
27	Gloa Tauro	1,449		•	
28	Manila	1,358		•	
29	Hampton R.	1,232		•	
30	Osaka	1,200		•	
31	Le Havre	1,185		•	
32	Genoa	1,180		•	
33	Tacoma	1,159		•	
34	Charleston	1,151	•		
35	Bangkok	1,099		•	
36	L.Chabang	1,036		•	
37	Melbourne	1,029		•	
38	Durban	984	•		
39	Barcelona	972		•	
40	Tianjin	935	•		

(table continues on following page)

Table 4.3 continued

Port	*TEUs thousand*	*Public*	*Mixed*	*Private*
41 Jeddah	921		•	
42 Southampton	891			•
43 Montreal	870		•	
44 Taichung	842		•	
45 Valencia	810		•	
46 Santos	829		•	
47 Sharjah	815		•	
48 Houston	798		•	
49 Sidney	765		•	
50 Miami	761		•	

Source: Cass (1996).

Privatization and Regulation of Seaports

During the past years, the world has observed a trend toward an increasing participation of the private sector in seaports. The traditional mode of port organization, with substantial public intervention, has become obsolete and unusable for adapting to the rapid changes in the industry. Due to unsatisfactory performance by public ports (high tariffs, inefficient services, overstaffing) and to tight fiscal constraints, many countries have opted for changing the legal frameworks for port operations and promoting the entry of private firms to invest in ports and to provide services.

Even if ports are privatized, the public sector must keep a role as a regulator of the activity of private operators. The regulatory activity of government in ports has two dimensions. The first one relates to safety, environmental issues, and quality of port services. The second dimension is the economic regulation of private port operators. As discussed below, the need for this type of regulation depends on the conditions of each port. While in some cases governments should keep control over tariffs and performance of port operators after privatization, in some others port services can possibly be provided under market conditions.

The Traditional Seaport Organization

Many different types of port organization can be found around the world (see table 4.4). Northern Europe uses a municipal model and southern

Table 4.4. Financing of Port Infrastructure in Different Countries

Country	*Maritime access infrastructure*	*Port area infrastructure*	*Port area superstructure*	*Land access infrastructure*
Argentina	P.A./Private	P.A./Private	Private	Most port authorities are responsible for road and other transport connections within port areas. Connection to the hinterland is usually the responsibility of governments. Regarding railways, responsibility can be national (Belgium), the port authority's (Germany), or the railway concessionaire's (Argentina). In the case of Hong Kong, China the private sector is responsible for infrastructures within the port area.
Belgium	State	Public	Private	
Cyprus	P.A.	P.A.	Concession	
Denmark	P.A.	P.A.	Private	
Finland	P.A.	P.A.	Private	
France	State/P.A.	Public/P.A.	Concession	
Germany	State	Public	Private	
Greece	State	Public/P.A.	Concession	
Hong Kong, (China)	P.A.	Private	Private	
Ireland	P.A.	P.A.	Concession	
Italy	State/P.A.	Public/P.A.	Concession	
Malta	State	P.A.	Concession	
Mexico	P.A.	P.A.	Private	
Netherlands	State	P.A.	Private	
Portugal	P.A.	P.A.	Concession	
Spain	P.A.	P.A.	Concession	
Sweden	P.A.	P.A.	Concession	
United Kingdom	P.A.	P.A.	Concession	
Venezuela	P.A.	P.A.	Private	

Note: 1. P.A.: port authority (financed with own resources).
2. Public: Financed by central, regional, or municipal governments.
3. Concession: In cases indicated, superstructure is publicly owned but operated by private firms.
4. This classification refers to the main seaports in each country. It is always possible that within the same country, ports with alternative modes of infrastructure financing not shown here may exist.
Source: European Parliament (1993); ESPO (1996).

Europe and South America follow a state model, in which governments control all main ports. A private model is when private firms or port authorities pay investment costs through charges on port users. This is more common in countries with a British tradition, where ports are regarded more as commercial rather than as public institutions.

Some point to examples illustrating a lack of relationship between seaport efficiency and type of ownership. A comparison between Singapore and Hong Kong (China) is often made: both are highly efficient ports; the former is completely public, and the latter private. Even though some exceptions may exist, however, finding a gap between private and public seaports is common in terms of efficiency.

One can usually characterize a traditional public seaport, before the introduction of reforms, by the following features:

- State or local government budgets are used to finance construction of large infrastructures, but these public budgets are becoming tighter.
- A public port authority finances the maintenance and repair of infrastructure.
- The port authority is financed partly with public funds and the rest by port tariffs and fees from private firms operating in the port.
- An excess of employees work in the port who have a high degree of unionization and strong positions at collective bargaining.
- Port efficiency in terms of tariffs and waiting times for ships is relatively poor.

This list of stylized facts does not reflect the exact situation of all public ports in the world, but merely pinpoints the main issues facing those seaports that have started introducing reforms. Other ports that are still run exclusively within the public sector will surely have to respond to similar challenges in the near future.

Some changes have occurred in the maritime transport industry in the past decades, which are irreversible and affect all world ports. These are basically technological innovations in the transport of cargo. The two most remarkable are the containerization of cargo and the development of large specialized ships.

These transformations have forced seaports to modernize their infrastructures and to buy new equipment capable of providing new services demanded by shipping companies. One can observe the trend toward containerization in figures for world traffic of cargo. The fast growth of the use of containers started in Europe, Japan, and the United States, but it is now spreading elsewhere. Table 4.5 presents figures for different regions of the world. The spectacular growth rates in Asia, South America, and India, and in general, all developing countries, is remarkable. Four of the five largest world ports in terms of container traffic are located in Southeast Asia: Hong Kong (China), Singapore, Kaohsiung (Taiwan, China), and Pusan (Republic of Korea).

Changes in the types of ships that maritime transport companies use become evident when studying the increase in the size of ships performing long-haul services. Economies of scale are evident in the use of larger,

Table 4.5. Containers: World Regional Traffic, Selected Years
(million TEUs)

Region	*1980*	*1986*	*1990*	*1995*	*Change 1980–95 (percent)*
Southeast Asia	9.08	19.10	32.42	61.84	581
Europe	11.49	17.76	23.25	33.06	188
North America	9.92	13.42	16.49	21.85	120
Caribbean/Central America	0.96	2.68	3.56	5.39	461
South America	0.38	1.04	1.44	2.76	626
Middle East	1.38	2.32	2.90	5.40	291
India	0.26	1.08	1.83	3.17	1,119
Australia	1.61	1.95	2.33	3.46	115
Africa	1.27	1.74	2.42	4.66	267
Total	36.35	61.09	86.64	141.59	290

Source: OSC (1996).

specialized ships, which allow them to transport larger volumes of cargo at lower unit costs. This has led to spectacular rises in the capacity of ships. At present, the latest generation of container ships, Post-Panamax, with a capacity of between 10,000 and 12,000 TEUs (20-foot equivalent units, a standard measure commonly used in the container industry), and a width of 20 containers abreast on deck, is now at the design stage (Hayuth and Hilling 1992). These huge ships will likely start operations in a few years, requiring drastic adaptations of handling equipment in terminals. For the transport of oil and other liquid cargo, most of the world's tanker fleet is now above 300,000 gross registered tons.

These two "revolutions" in the maritime transport industry have forced seaports to start rapid facilities renovations to serve the new needs of shipping companies, especially the growing demand for container handling services. Seaports now face a more competitive situation than in the past, and so they must have the required facilities as well as low prices, or they risk losing traffic to rival ports. However, shipping companies are increasingly working with hub-and-spoke networks; therefore, they demand the services of large ports that act as connection nodes (hubs) where cargo is transferred to smaller ships that perform regional-based services (feeder services).

To make the necessary investments to meet the growing demand, the seaport industry has strong capital needs. Some large ports have the possibility of becoming hub ports at a regional level, and thus attract large volumes of traffic. The following is a list of the challenges that this situation poses for modern ports, especially in the context of reduced public subsidies, which are due to tighter fiscal conditions among governments.

- The need to seek financing for infrastructure renovation and building of new facilities
- The need to achieve high efficiency levels in costs and operation times and to keep prices low
- The reduction of excess of labor, aggravated by the trend toward intensive use of capital at seaports.

Thompson and Budin (1997) have identified several reasons for introducing private participation in transport industries through the use of concessions. First, the private sector can provide services at lower costs than the public sector, because it usually is more productive and efficient. Second, if private capital is used to finance costs, the public sector can devote its scarce resources to other priorities. Last, the private sector is generally more able to search for business opportunities and to respond more swiftly than the public sector to changing conditions in competitive markets.

Introducing private participation in seaports appears to be a response to the challenges pinpointed above. This is a worldwide trend: ports in general are adopting the landlord model. As described earlier, this model allows port authorities to retain ownership of the infrastructure to avoid the risk of monopolization of essential assets by private firms. The private sector then operates these assets. Concession contracts between port authorities and private firms are the most common instrument to allow private participation in ports.

The role of port authorities is thus transformed from their traditional concept as institutions in charge of all port activities (see table 4.3) to a new one in which they only coordinate these activities. When introducing private firms into seaports, designing regulation systems to monitor the behavior of private operators is a new need. This regulation would usually take place in asymmetric information conditions (firms know their costs and market conditions better than the regulator does). Port authorities are not strictly required to perform this regulatory role at seaports, however. Instead, an independent institution could perform this task.

The Movement toward Privatization

In the choice of best form for introducing private participation in the organization of port services, several alternatives are possible, depending on port size, initial conditions, and type of service. These include the following:

- *Full privatization:* Selling the seaport as a whole. Justified by serious fiscal needs in the public sector, this method transfers all assets and liabilities to the private sector.
- *Build-operate-own:* Transferring parts of the seaport to private operators for development. Short-term financial needs justify the use of this form of privatization.
- *Build/rehabilitate-operate-transfer:* Introducing private participation in the port to build or renovate facilities required for service provision. In this case, the public sector does not lose ownership of the port infrastructure, and new facilities built by private firms are transferred to the public sector after a specified period of time. This is the case of classic concessions.
- *Joint ventures:* Creating a new, independent company by combining the efforts of two or more firms. This type of agreement arises when two parties with common interests join forces. For example, in some cases, one firm supplies technology and know-how, while another has knowledge of market opportunities and customer contacts.
- *Joint ventures not exclusively between private firms*: Creating collaborations between, for example, port authorities and private firms, as in the cases of Shanghai (China), Kelang (Malaysia), and other Asian ports with large investment projects, where port authorities have formed many joint ventures to develop and operate new terminals. In other cases, collaboration may be between several public firms, as in the Singapore PSA Corporation with the port authority of Dalian to develop and operate a container terminal in the port of Dayaowan (China). However, a port authority's participation in joint ventures regarding projects within the port that it regulates raises some concerns. The most important is that the port authority has to regulate and at the same time be a part of the regulated firm. The joint ventures between port authorities and private firms should then be only a temporary solution to promote development when private participation is lacking, but port authorities must return quickly to their primary role as regulators.
- *Leasing*: Port authorities simply renting port assets to private operators for a fixed period to obtain income from contract fees. Contrary

to concession contracts, firms that lease are usually not required to make investments, therefore they only assume commercial risk. Port facilities such as storage buildings and cranes are frequently leased to operators.

- *Licensing*: Port authorities allowing private operators to use their own equipment to provide some services for which the required equipment is relatively simple. Private operators generally own these assets, and the infrastructure (as well as some superstructure elements owned by the port authority in some cases) is made available to them at a specified fee. Stevedoring companies, pilots, tug operators, and consignees can work under this type of agreement.
- *Management contract:* Introducing private participation in a port in a simple form by contracting out the port management. In this situation, the port authority is the owner of the infrastructure and port facilities, but a private firm can provide a more commercial approach to operations. The public sector in this case faces both investment and commercial risks, because managers do not invest their own capital in the port. The port of Bristol (United Kingdom) is an example of this type of contract, where facilities are owned by the local government, but the port is privately managed.

When looking at the foregoing options to determine which is the best alternative for a particular port, one must evaluate the port's objectives and consider the constraints that the port authority faces. The type of service may determine the possible degree of private participation. A basic determinant would be to consider whether or not the service requires the exclusive use of a port's fixed assets.

SERVICES THAT DO NOT REQUIRE EXCLUSIVE USE OF INFRASTRUCTURE OR SUPERSTRUCTURE PORT FACILITIES. Within this group are services such as pilotage, towing, consignees, and other ancillary services to ships and crews. In many ports, as a result of safety arguments, the public sector traditionally provides some of these services. In particular, berthing services are usually considered a public service obligation—in other words, every port user has a right to be provided with these services—and port authorities therefore directly provide them to avoid the possibility of disruption of service.

Pilotage is a typical example of a compulsory service organized on a monopolistic basis in many ports. Pilotage is required for ships above a given capacity or length, and for dangerous cargo. Most ports have exemptions, such as for regular passenger services (ferries). The degree of public intervention in pilotage varies across countries. In some, pilots are civil servants. In other cases, they are organized as independent agents, basically self-regulated by their own associations.

Regarding other berthing services—towage and tying—diverse solutions can be found among ports. Both are generally considered public service obligations, and port authorities perform them directly or grant licenses to private operators. A single firm can provide towing services exclusively, or in large ports, having several competing companies (De Rus and others 1995) is feasible. Some ports do not strictly control these services, only setting minimum standards (technical capacity, safety and environmental standards, and so forth) for private operators to obtain licenses. In this case, market conditions determine towage tariffs, which are not fixed by the port authority.

Therefore, one can conclude that this set of auxiliary services to ships can work reasonably well through a system of licenses by which several operators are authorized to provide services within the port area. These operators' activities can be regulated in terms of prices and quality of service. In some cases, having several competing operators (such as consignees or pilots) is possible. Then, strict price regulation would not be required, unless collusive practices are detected.

For other services, such as towage, having several operators can be complicated, because they are limited by port size. In the case of medium or small ports, establishing limits on prices and service conditions is needed to avoid market domination by a few firms that could exploit their position to extract rents from port users.

SERVICES THAT REQUIRE EXCLUSIVE USE OF ASSETS. These services require the use of one of the most scarce resources at seaports: space. They include terminals for cargo handling, storage areas, repair docks, and fuel suppliers. Introducing private participation in these services is more complicated, because operators need to use assets that are considered to be optimally owned by the port authority. Therefore, concession contracts need to be written carefully to reconcile private operators' interests with port authorities' objectives. At the same time, contracts must include incentives for private operators to maintain or enhance assets as required.

The number of operators for these services is by definition extremely limited, although it will vary according to port size. Similarly, the need to establish some price and quality regulation depends on what type of port it is and how many alternative ports are available in the area. For example, in the case of a port in a highly competitive region, the port authority or the institution in charge of regulation does not need to be extremely concerned about excessively high charges by private operators. In that situation, private firms self-regulate prices to avoid the risk of losing their market share to competitors.

The Need for Port Regulation

Box 4.1 presents a possible characterization of the different sizes and degrees of development that a seaport can reach. Depending on port size, having economic regulation of ports where private operators use fixed assets is more or less required. We can broadly distinguish two types of situations according to the degree of development reached by a seaport. First are those ports with a reduced market size—port types 1 and 2—that do not require more than a general cargo terminal, which can serve all kinds of goods and containers, or they possibly have one terminal specializing in dry bulk goods.

For these small ports, one can consider introducing some form of competition among those firms that are willing to operate in the port. Thus, establishing a system of auctions is possible in which private firms bid for

Box 4.1. *Levels of Port Development*

1. *Small local ports*: These serve small communities; therefore all kinds of general cargo and containers pass through them, usually transported by relatively small ships (short sea shipping services). The basic facilities are general-use berths with storage areas nearby.
2. *Large local ports:* As traffic reaches a certain level, investing in specific equipment is profitable, such as a dry-bulk terminal with berths to serve deep-draught ships. Some investments are probably also made to improve land access and buy container-handling equipment, although general cargo berths would still be used.
3. *Large regional ports:* A seaport that handles a significant level of long-haul traffic requires large investments in specialized terminals, such as container terminals and facilities for specific goods (coal, oil, grains, and so forth). These ports can serve huge ships, more than 60,000 gross registered tons, used in long-haul bulk transport.
4. *Regional distribution centers:* The world's largest ports—Rotterdam, Hong Kong (China), and Singapore—are collections of highly specialized terminals that only serve particular traffic. They have excellent equipment for transport interchange among all modes (railways, road, inland navigation). Their role is to act as hubs for huge long-haul ships to conduct transshipment operations. From the hubs, smaller ships or other transport modes distribute cargo to the region.

Source: Stopford (1997).

the right to operate the terminal. Once the bidding process is over and a single operator is chosen, having some regulation over the charges that this firm imposes on port users is necessary, because it would otherwise enjoy a monopoly position. Price-cap systems or rate of return regulation are alternative options for regulating the behavior of private operators, depending on the information and experience that the regulatory institution has in the type of service subject to regulation.

This need for regulation, however, is less strict if competition exists among ports. In cases in which a region offers alternative ports for shipping companies, regulating prices is less necessary, because the market mechanism makes the private operator either keep prices low or risk losing traffic. If alternative ports do not exist or do not have adequate facilities, the private operator enjoys market power that must be controlled by regulation. For example, users of the Mexican ports of Veracruz and Manzanillo complain about high tariffs, arguing that operators enjoy market power. Although alternative ports exist both along the Atlantic and the Pacific coasts of Mexico, the antitrust institution of the country has imposed regulation on prices for the operators of container terminals at both ports.

In the case of large seaports—types 3 and 4—the volume of traffic is large enough (for example, more than 100,000 TEUs) to allow competition to exist within the port. If a large port is divided into several independent terminals, inducing competition between operators is possible for traffic that calls at the port. Here, price regulation is less of an issue, because if the market mechanism works reasonably well, private operators will restrain price increases themselves. Some form of supervision is still needed, however, because the situation is prone to collusion between competitors (due to the small number of parties involved).

As an example, Argentina recently split the port of Buenos Aires into five different concessions for independent companies to operate its terminals. Although some problems occurred initially, and the number of operators has been reduced, improvements in port outcomes have been substantial. Worker productivity rose from 800 tons in 1990 to 3,100 in 1997, and waiting time for containers was reduced from 2.5 to 1.3 days over the same period.

Therefore, this analysis concludes that introducing private participation in the seaport industry appears to be the most attractive option for ports trying to develop and adapt to the new conditions of the maritime transport market. Modern ports are in need of huge investments to enhance their facilities, to be able to provide the services that shipping companies demand. Because the financing of required investments is increasingly difficult for governments in all countries, the optimal solution is to try to attract private capital for investment in ports, and to improve efficiency through the liberalization of port practices and the introduction of competition. The role of public sector institutions then changes from direct provider of services to regulator and supervisor.

The cornerstone of port systems is now the correct design of concession contracts for collaboration between the public and private sectors. As discussed above, the existence of competition determines the need for the regulation of private operators, but even when competition is present and regulation is not required port authorities still need to have some degree of control over the infrastructure assets that private firms are using.

Figures in table 4.6 indicate the feasible degrees of competition between operators in different situations so as to assess when regulation is required. Establishing universal threshold values for all ports and types of cargo is difficult, except for containers, on which experts seem to come to a consensus.

These figures show that if the volume of container traffic in a port is less than 30,000 TEUs per year, having several terminals and operators does not make sense, because the market is extremely small. The best solution is having a single operator and regulating its charges. If traffic is greater than 30,000 TEUs but less than 100,000, having several operators, possibly sharing a single terminal, is feasible. This would be a situation of intraterminal competition, with cargo handling services provided to port users by various stevedoring companies that would make use of the equipment (cranes) the port authority owns, or they would employ their own equipment, depending on their financial positions.

If traffic is greater than 100,000 TEUs, the port has the possibility of opening different terminals, which several companies can operate; they make use of separate berths and manage them better. Competition is easily implemented between terminals. When a company serves all ships using a given berth, port authorities also can make the private operator responsible for collecting port

Table 4.6. *Threshold Values to Determine the Type of Competition: Container Traffic*

Type of competition	*Level of traffic (TEUs)*
Intraterminal	30,000
Interterminal	100,000
Interport	300,000

Source: Kent and Hochstein (1998).

tariffs from users (charges for the general use of the port, different from the prices the operator charges in the concept of cargo handling) and transferring revenues to the port authority. In this range of traffic volume, providing incentives for private operators is also possible to finance projects for infrastructure enhancement, or even for the building of new facilities.

Finally, in a region where container traffic is greater than 300,000 TEUs per year, the market allows for several alternative ports to exist and compete for traffic. Interport competition is likely, which reduces the need for control over private operators' prices. Even in this case, however, paying attention to the drafting of concession contracts is necessary, because private operators must be compelled to fulfill their obligations not only regarding service conditions and charges, but also regarding equipment maintenance, safety, service quality, and all other matters that are costly for the concessionaire and could be underprovided.

Concession Contracts in Ports

The different seaport services that private operators can offer are subject to a variety of possible contracts. As discussed above, port size is one of the key variables in determining the type of private participation that one can choose. For those ports with insufficient demand to allow for the existence of several terminals, the best idea is probably to transfer the port as a whole to the private sector. If desired, keeping public ownership of the infrastructure would be feasible, but a single operator could run the port, providing both the infrastructure and cargo handling services (stevedoring, storage, and so forth). The port operator could also provide the rest of the complementary services (such as berthing), or if demand is sufficient, it could be open to competition among different firms.

In the case of larger ports (the landlord type), introducing private participation in more sophisticated forms is feasible. It is in these ports that infrastructure can be split into separate terminals, generating competition within the port. For services that are easy to specify in a contract and do not involve the use of substantial infrastructure, private firms operating under licenses could provide them. For other port services that require the exclusive use of infrastructure or superstructure, private operators must be subject to concession contracts, in which conditions are stipulated to determine when a private operator can use assets, and what its obligations are.

One can specify license contracts relatively easily, because in general, the operator owns the equipment required to provide the service. The role of the port authority or any other regulating institution is limited to imposing minimum standards (for example, professional qualification for pilots, or number and power of tugs for towing companies) and establishing some rules for service provision.

A concession contract is, by definition, more complex than a license, because it involves not only questions about service provision, but also about adequate maintenance of assets, investments to be made, and risk allocation between the regulator and concessionaire. All these aspects are discussed in detail later. Concession contracts can be regarded as an intermediate solution between public ownership and full privatization of a port. Private participation is introduced to achieve efficiency gains in the industry, and at the same time political concerns are safeguarded by allowing society to retain ownership of essential assets (Crampes and Estache 1997). Other industries involving expensive infrastructures (electricity, water, gas) have extensively used these contracts, for collaboration between public and private sectors.

When designing a concession contract, one must carefully tailor several aspects: the object of the concession, exclusivity in the use of assets, the concessionaire's obligations and payments, the term of the concession, penalties and fines, and risk allocation (Crampes and Estache 1997; Kerf and others 1998; Thompson and Budin 1997). When writing concession contracts, one also must consider the problem of excess of labor, common to almost all ports worldwide. Another relevant feature is to carefully design the selection method to determine the winner of the concession.

OBJECT OF CONCESSION. The first question to be answered when drafting a concession contract is what is to be concessioned. Even though this may sound like a simple question, the contract must be precise about the assets that are to be transferred to the concessionaire, the services that it must provide, the services that will be left to the public sector, and those other services that are subject to open competition. Thus, in the case of the concession contract for a terminal, the contract must describe in detail the limits between the infrastructure that is concessioned (berths, surfaces, inner access roads, and so forth) and what is not (such as general roads for intraport connections), to clearly establish the concessionaire's responsibilities.

The port authority, or the regulatory institution signing the contract, must guarantee that the assets are transferred to the concessionaire free from any other contractual obligations, and that they are available for the agreed terms and times. Avoiding delays in the transfer of assets is important, particularly those owed to the negligence of the port authority in fulfilling the terms of the contract, because this may impede the concessionaire from promptly starting its operations.

A concession contract must explicitly define those services to be provided by the concessionaire, and on what terms. For example, the contract prepared for the concession of terminals at Puerto Nuevo in Buenos Aires established that the concessionaire is the exclusive provider of the following services:

- Reception, delivery, stevedoring, and storage of cargo
- Administrative control of loaded and unloaded cargo
- Safe berthing and unberthing of ships
- Any other service to ships or cargo promoting efficiency of the terminal.

In this example, the ambiguity of the last point is remarkable, because it opens a door for a concessionaire to interpret what has or has not been included, and thus to claim for itself the exclusive provision of some service that, in principle, the port authority did not plan to concede. One should carefully avoid this type of ambiguity in concession contracts to avoid litigation problems with private operators.

Exclusivity. A concession contract must specify what services the concessionaire is to provide exclusively, and what other services are open to other firms. For example, in the case of the concession contract of the port of Mar del Plata (Argentina), which is a small port concessioned as a whole to a single company, it explicitly states that the berthing and other complementary services to ships (energy and water supply) should be regarded as public services. The concessionaire does not have the right to exclusively provide those services. It could offer the services to its customers, but it must also allow access to any other interested company. Conversely, some services are left exclusively to the port operator (cargo handling, marketing, and so forth).

Regarding this question of exclusivity, the rule should be to guarantee port users an efficient provision of services. For services in which the presence of a competitor could be positive, one should allow access to any interested party. Meanwhile, for those activities in which competition can result in a deterioration of services for users, exclusivity is desirable. For example, consider the case of a container terminal that is concessioned to a private operator, but other operators are allowed to enter to handle general cargo with their own cranes. Even though competition in the general cargo segment could improve, avoiding interference with the container operations by allowing a single operator to exclusively perform all cargo loading services (general and container) is probably preferable.

For services that involve some public service obligation (providing desirable but nonprofitable services), having a single firm provide them exclusively is usually simpler. If several operators are forced to offer these services, subsidies must be paid to all of them, which will probably raise the administrative costs of controlling the system.

In small ports, concessionaires must be offered guarantees that they will be able to recover their investment costs. To do that, one commonly includes provisions for the concessionaire to obtain some compensation in the event that during the contract term, another facility is built within the port and it reduces the traffic level the concessionaire expects. These provisions may take the form of minimum traffic guarantees or priority for the concessionaire in the bidding process for the building of new facilities.

Obligations and Payments. A concession contract must explicitly mention the concessionaire's obligations in terms of level and quality of service. It also should clearly specify how charges to users are to be determined, who owns the revenue obtained from those charges, and what payments are to be made between the parties.

The norm is for the concessionaire to pay a fixed annual fee (sometimes called a canon) to the port authority or the institution responsible for the concession. The contract can be designed with a negative fee. In that case, the concessionaire receives a payment from the port authority, as a payment for public service obligations, if revenues from port users do not cover costs.

Concessionaires' fees can sometimes be linked to the level of traffic served by the terminal or the infrastructure subject to concession (for example, making it proportional to tons or TEUs handled), including a guaranteed minimum payment to the port authority. Concessions for container terminals granted in the port of Buenos Aires and in Brazilian ports have used this system of fees proportional to port activity.

In some cases, the concessionaire can be made responsible for collecting port dues charged on ships and cargo for the general use of the port and transferring the revenue to the port authority. When performing that task, the concessionaire acts as an agent for the port authority, because the level of port dues is determined by the latter. Meanwhile, the level of charges for services the concessionaire provides (cargo handling, storage, and so forth) is usually freely determined, although it is subject to some form of external regulation. As discussed in the previous section, the need for regulation is more important for small ports with a single terminal, but it is not so relevant if intraport or interport competition exists.

Concession contracts in the seaport industry are usually associated with building or rehabilitating facilities (build/rehabilitate-operate-transfer type of contracts). In those cases, the contract must clearly specify the starting and completion dates of the operation, as well as the moment when ownership of the assets is transferred to the port authority. Technical issues about infrastructure building (materials, methods, and so forth) also should not be left to the concessionaire; instead, the concession documents should specify them in detail.

TERM. No universal rules about the proper length of a concession exist. Economic theory on regulation indicates that the longer the life span, the more incentives the concessionaire has to make adequate investments to enhance assets, because profitability depends on the state of the facilities. The longer the period between two concessions, however, the less information the regulator has on cost and demand conditions. Therefore, a trade-off exists between incentives and information for optimal regulation of a concession.

In addition, concessions associated with large investments must allow sufficient periods for operators to recover construction costs. In practice, concessions with large projects are usually longer than those with no investment requirements. The average term of a concession contract is more than 15 years, and those with large projects can be around 25 years. Sometimes, provisions also allow operators to obtain the automatic renewal of concessions if they fulfill investment or rehabilitation plans. Table 4.7 shows examples of the concession terms for some contracts signed in ports around the world.

EXCESS OF LABOR. One of the common problems shared by many ports worldwide is an excessive number of port workers, generated partly by unionization and partly by the technological changes introduced in cargo handling procedures. In a relatively short period, seaport activities have been transformed from being labor-intensive to being capital-intensive, a process that has made a large number of employees redundant. The traditional public organization of seaports has exacerbated this problem, because port workers are civil servants in some countries and therefore have certain rights, in general, enjoying significant social benefits that must be respected.

The transformation of a port has to deal with this problem, because worker resistance can block any reform. Port workers can have significant political influence in some countries, Brazil, for example. Solutions include

Table 4.7. Term of Port Concession Contracts in Practice

Port	*Period (years)*
Buenos Aires terminals 1–4 and 6 (Argentina)	25
Buenos Aires terminal 5 (Argentina)	18
Mar del Plata (Argentina)	15
Manzanillo (Panama)	20
Karachi (Pakistan)	20
Le Havre (France)	50
Kelang-Westport (Malaysia)	30
Manila-South Harbor (Philippines)	15
Santos (Brazil)	25
Maputo (Mozambique)	15

Source: Authors.

providing funds to offer redundancy compensations and anticipating retirement schemes for workers. Generally, public budgets partly finance these funds, but private operators are also required to share the financial burden resulting from the labor problem.

Different countries have searched for different solutions to this problem in the concession contracts for their terminals. For example, in Panama, the port authority has offered to unions to employ a fraction of the income generated by concession fees to redundancy programs. In Mexico, the government and unions reached an agreement in which terminal operators have a right to negotiate only with the workers they employ, instead of forcing them to bargain with a single centralized port union (Brennan 1995).

Penalties and Fines. To guarantee adequate compliance with the terms of a concession agreement, the contract must specify a series of penalties and fines that the concessionaire must pay to the port authority in the event of default. For example, if a private operator does not fulfill its obligations in terms of investment requirements or quality of service, the port authority might have the ability to raise the concessionaire's fixed annual fees.

For this mechanism to be valid, the port authority must regularly perform some inspection tasks to verify if the concessionaire is providing the required services and keeping assets in the stipulated conditions. If a port authority is excessively permissive with defaults from the concession, it could damage its reputation with other operators for future concessions. Because concession contracts have long lives, port authorities should establish strong positions from the start of the concession. In addition, the imposition of sanctions on one operator will usually have a demonstrative effect on other operators within the same port, or in other ports regulated by the same institution.

RISK ALLOCATION. One of the more complicated aspects of designing a concession contract is the adequate allocation of risks. As in any other sector in which this type of contract is being implemented, the ideal rule is to allocate each type of risk to the party that can take better actions to avoid it, so that all agents are provided with incentives to behave optimally. According to Kerf and others (1998), the following types of risk are involved in a concession contract:

- *Design/construction risk:* This type of risk appears in contracts that require the concessionaire to make investments on a building or rehabilitation project. Once construction starts, the concessionaire may try to renegotiate the contract, arguing that unforeseen circumstances have arisen or costs were incorrectly estimated. The concessionaire tries to obtain financing from the public sector to proceed with the project, or at least a reduction in the fees that it pays to the port authority.

  In this case, the advice is to study the origin of the forecasting errors. When errors can be attributed to defective information or mistakes in the bidding process granting the concession, the government or the institution responsible for ports should assume responsibility and pay the concessionaire for the extra costs. A completely different situation arises if cost deviations are caused by poor estimates on the part of the concessionaire. In that case, the port authority's position should be strong, to make the operator cover the extra cost plus any penalty established in the contract if the building is not completed by the stipulated dates.

  For investments in new, commonly built infrastructures, such as container terminals, international standard designs allow for the estimation of reasonable costs and completion periods for berths of a given length and width, providing that normal subsoil conditions exist (DSC 1998). Therefore, this type of risk is relatively low for standard investments and should be allocated to builders. A different situation arises if geographic conditions are not standard, or a project has special characteristics, in which case the builder can be allowed some margin of error.

- *Operating cost risk:* Another source of risk is the existence of higher than foreseen costs for providing service. Again, the concessionaire should assume all excesses of costs that can be reasonably predicted. If the bidding process was correctly designed, all bidders had the same information, and therefore they must have carefully devised their cost estimations. With that in mind, the bidding process must arrange for all candidates to have permits to inspect the involved infrastructure and receive as much information as required. If this provision is made, any excess costs discovered thereafter should be the concessionaire's responsibility.

  The possibility that the port authority can cause some excess costs must also be considered, however, in which case the concessionaire should be allowed to renegotiate the contract or be compensated accordingly. For example, some cost rises could be due to delays in obtaining required permits, terminating existing contracts with other firms that have rights over elements included in the concession (such as occupied buildings), or disposing of obsolete assets that the port authority had agreed to remove. In all of these cases, if the delays the port authority imposed on the concessionaire result in losses or higher costs, the latter should be able to receive compensation.
- *Revenue risk:* This is one of the more dangerous risks in the seaport industry, as in any other sector in which the concession contract is valid for a long period. If demand forecasts used to compute the expected income flows are too optimistic, the concessionaire could eventually end up with much lower revenue than expected, and could even go bankrupt. The general rule for this type of risk is to allocate it to the concessionaire, to provide incentives for candidates to properly estimate the expected demand levels in the bidding process. Furthermore, if regulations on charges allow the operator to lower them freely, the concessionaire could minimize this risk by reducing charges to attract more traffic during a period of low demand.

  A strict application of this rule, however, implies that in some cases, an operator should be allowed to go bankrupt, because otherwise the system would lose its credibility. In that situation, the port authority must consider what the options are after the private operator ceases its service provision. In large ports, other operators could supply those services, until a new bidding process is launched, but in the case of small ports, the port authority must be able to provide services directly or quickly replace the operator, or the port will suffer a long period of inactivity.

Another situation that one should carefully study is whether the risk could be mitigated in some low-demand situations by introducing flexibility in the regulation systems. In some cases, price limits imposed on concessionaires could result in low revenue flows to the operator, not allowing it to adequately recover its costs. If that is the case, and the operator is proven to be not negligent in letting its costs rise excessively, the regulator should be more flexible and rescue the financially strained concessionaire.

- *Financial risks:* In developing countries, currency values are usually subject to wide oscillations. Therefore, all seaport projects implemented in these countries are subject to exchange rate risks, especially for those with longer terms. Relatively easy solutions can reduce this risk, however, such as nominating all monetary references of the contract in a hard currency, or buying insurance to cover it. Similarly, interest rates can also suffer large variations that could substantially alter operational or building costs. Even though both parties should privately cover this, concession contracts may also include provisions on financial risk.
- *Environmental risks:* Some of the circumstances that one has to consider when drafting a concession contract include accidents within a port area or in its access zones, which can have disastrous effects for the port and adjacent areas (for example, oil spills). For private operators to minimize those risks, they should be strictly liable for any accident caused by negligence in maintaining adequate signaling devices or in not fulfilling required operations such as dredging. Even though the port authority should have subsidiary responsibility in compensating affected parties for those costs not covered by the concessionaire, it should supervise private operators to ensure that they are properly insured to cover their civil responsibilities.

  In the construction phase, the concessionaire must be strictly supervised so that it takes care of any negative environmental effects that it might cause, such as dumping of dredging materials or impacting adjacent areas. The contract should explicitly include these aspects to ensure the correct allocation of responsibility.

Selection process. One of the more important elements for a concession contract is carefully designing the selection process, deciding which firm or consortium will be granted the concession. This process must pursue the objective of determining, in conditions of asymmetric information, which candidate can more efficiently run the assets that are the object of the concession, and whether that candidate has the ability to implement the associated building/rehabilitation projects.

The usual practice is to design a selection process based on two consecutive phases as follows:

- *Prequalification*: In the first stage, those firms satisfying several criteria are selected to be evaluated on their proposals. Criteria used for prequalification include experience in the seaport industry and a minimum financial capacity. This prequalification reduces the candidates to a small number, so their proposals for the project can be studied in great detail. Generally, candidates present the information required for this first stage separately from the economic proposal (usually enclosed in different envelopes), and it is evaluated objectively. For example, in the case of the concession for the port of Mar del Plata (Argentina), the information about experience and financial capacity was condensed into a single index, and only those consortia above a certain minimum value qualified for the next stage.
- *Concession award*: In the second stage, the proposal that is closest to the objectives the port authority is pursuing is selected. Thus, in general, the winning proposal is the one that offers a higher fee payment to the port authority (if financial need is the basic reason behind concessioning port assets) or the one that offers the lowest charges to port users (if port efficiency is primarily sought). If the concession involves investment projects, including an evaluation of the best project, or the one that has the lower cost, is also possible. In the example of Mar del Plata, the solution was to summarize all three of these criteria on a single index. The final decision was then based on the information from this index, to which the first index calculated in the prequalification stage was added. This example constitutes a sophisticated system of firm selection, because it uses all of the information provided by candidates.

In summary, one must design the process of selecting a concessionaire according to the objectives of the government or the port authority. The process should be as transparent as possible, and try to avoid allowing candidates to collude on their bids. In the case of the concessions for the terminals of the port of Buenos Aires, the selection process was designed so that a candidate was only allowed to win one terminal, as a way to promote competition in the bidding process (not all terminals had the same characteristics, and therefore all bidders were supposed to tailor the bids to win the more attractive assets).

RENEGOTIATION OF CONCESSION CONTRACTS. Finally, a relevant aspect that one must consider when drafting a concession contract is that, in all probability, during the life of the contract some unpredicted circumstances will arise and force parties to renegotiate. This statement is true for any kind of contract, but it is especially important in the case of concessions. This is due first to the long period of the contract, which makes anticipating all possible contingencies unfeasible for any party. Unforeseen contingencies also occur because concession contracts for port facilities are related to expensive fixed assets that cannot be easily removed and redeployed in another location. One must study renegotiation, because in the case of disagreement, the owner of the asset is the agent that has rights over its use. If renegotiation is not anticipated, the port authority can find itself in a weak position, allowing concessionaires to extract ex post additional rents.

For example, consider the case of a small port that is concessioned as a whole to a single operator. After the concession is granted and operations start, the concessionaire tries to renegotiate the contract to obtain better conditions by using the threat of stopping the provision of services to ships and blocking the use of assets by an alternative operator. If the concession contract does not precisely define who owns the assets and when they can be transferred between parties, the concessionaire could claim valid rights over the assets granted by the concession, and litigate against the port authority (hold-up problem). If the contract clearly states that the port authority owns the assets, however, in the case of renegotiation, the concessionaire would never be in such a strong position, because, as owner, the port authority can always "rescue" the assets and keep the port working.

Renegotiation of a concession contract is probably the rule and not the exception, and one should not perceive it as a failure. Because concession contracts are typically long-life documents, the parties could not foresee all possible future contingencies at the moment of drafting the contract. Knowing this in advance, the parties should consider several future conflict scenarios and ensure that some provisions are included to establish at least basic renegotiation rules. Nevertheless, a concessionaire should try to avoid renegotiation at early stages as much as possible, because the credibility, transparency, and fairness of the bidding process can be put in jeopardy.

International Experiences

In general terms, the process of privatization and liberalization of the main seaports worldwide has been characterized by the use of concession contracts, rather than by selling seaport assets to the private sector. Through concessioning, port authorities reduce their functions and are transformed into landlords responsible for coordinating all activities performed at the port. Consequently, they receive all rents accruing from asset renting.

The seaport industry is experiencing several patterns of privatization and deregulation, depending on the region and the initial situation of the ports. In Europe, the model increasingly seems to be the introduction of private firms in the provision of port services, but in general, infrastructure is kept within the public sector, and in some countries, governments continue to finance investment costs. A debate is going on within the European Union, however, on a plan to create a self-financing port system that would not receive subsidies from governments (European Commission 1997). The idea is that port authorities should adequately design port tariffs to finance infrastructure expansion or else seek private participation. In this region, limitations on the use of public funds are not motivated by the states' lack of financial capacity, but by competitive considerations. Some argue that in order to have a single market in which all ports compete on equal terms, having some states subsidizing ports, with other seaports self-financing their facilities, is not a fair scenario.

The United Kingdom has introduced the most radical reform in Europe, by fully privatizing most of its ports. In 1996, private institutions handled around 70 percent of all cargo. The process started by privatizing the Associated British Ports, an institution that had owned all former state ports. Subsequently, ports under a different legal status (trust ports) were transferred to the private sector. Results seem to be positive; investment figures have risen and private operators are making substantial profits (Ferrer 1997). Some critics have pointed out, however, that the British experience does not clearly benefit port users in the long run. Possibly, port authorities and regulatory bodies will again be required.

In Eastern Europe, the transition economies have some ongoing reforms of seaport systems, aimed at introducing private participation. Some ports have been transformed into state companies with worker participation, such as the St. Petersburg port, where employees own 51 percent of the port. Another example is the container terminal in the port of Vostochnyy (Russia), which is operated by P&O Australia, SeaLand, and a local Russian firm. Another example is the case of Polish ports, where a new Port Law, enacted in 1997, established port authorities as joint-stock companies, with 51 percent of shares remaining in public hands (state, treasury, local governments and municipalities) and 49 percent open to private participation.

In some Asian countries, private participation in financing infrastructure building began long ago in Japan and Hong Kong (China), where private firms build and operate infrastructure under long-term concessions. In other Asian countries (China, Korea, Malaysia, and the Philippines), shipping companies such as Maersk and P&O are also actively participating in the development of seaports.

Latin America is one of the more dynamic regions in terms of seaport concessioning, the building of new facilities, and most importantly, the rehabilitation of existing ports. The model of mixed public/private financing of seaports is successful in this region for three reasons. First, seeking capital is strongly needed to finance investments, because most governments have highly constrained budgets because of debt service payments. Second, rapid economic growth is generating new traffic that demands new facilities and more efficient services. Third, fierce competition makes upgrading of port facilities necessary or the ports risk being displaced by rivals. The experience of three countries in the region (Chile, Argentina, and Brazil) is discussed in some detail, while a summary of other experiences for Latin American countries and other regions of the world is presented later.

The Case of Chile. Chilean international trade is served by 38 seaports, of which 11 are publicly owned and organized by the public agency Empresa Portuaria Chilena (Emporchi), while 27 belong to the private sector. Of the latter, 11 are private (owned by mining and other companies), and 16 are privately owned but publicly used. Table 4.8 shows the distribution of cargo between ports. In 1997, Emporchi handled 94 percent of container cargo, 69 percent of general cargo, 18 percent of dry-bulk cargo, and 11 percent of liquid-bulk cargo. In total volume, the 11 public ports handled 37 percent of total tons that passed through Chilean ports (Tortello 1998).

In 1978, seaports in Chile were characterized by the split of cargo handling between two different groups of workers. Specialized port workers performed stevedoring operations, while Emporchi employees did loading and unloading operations. Both groups enjoyed some monopolistic positions. Stevedores had strong limitations on increasing their numbers, because each worker was required to have a special license *(matricula)* to be able to work as a stevedore. This practice transformed stevedores into

Table 4.8. Chilean Ports: Distribution of Cargo by Port Type

Port type	Containerized cargo		General cargo		Dry-bulk cargo		Liquid-bulk cargo		Total cargo	
	Tons	Percent	Tons	Percent	Tons	Percent	Tons	Percent	Tons	Percent
Private[a]	1,702	0	194,501	2	12,394,187	49	6,870,439	75	19,460,829	38
Private[b]	427,347	6	2,900,055	29	8,247,183	33	1,334,878	14	12,909,463	25
Emporchi	6,375,130	94	6,804,283	69	4,481,230	18	987,764	11	18,648,407	37
Total	6,804,179	100	9,898,839	100	25,122,600	100	9,193,081	100	51,018,699	100

a. Private ports for exclusive use by owners.
b. Private ports for public use.
Source: Tortello (1998).

monopolists for those services, which resulted in high tariffs and low productivity. Emporchi was by definition a public monopoly, working at the state level, and its workers constituted an important pressure group.

In 1980 the government decided to change the status quo. It introduced legal changes in 1981 by passing the new Seaport General Law, which effectively eliminated the monopoly of Emporchi in cargo-loading operations, allowing private participation in those services. Almost simultaneously, a different law abolished the system of licenses for stevedoring, allowing any worker to perform those services for shipping companies. The state made payments to compensate the 2,700 workers who lost their privileges and who were clearly opposed to any reform.

These regulatory changes permitted the significant entry of new private operators, establishing a competitive market for cargo handling. The impact on costs was substantial: cost savings of US$17.7 per handled ton for general cargo and US$1.43 per ton for dry-bulk goods, though liquid cargo had no improvement (Tortello 1998). The different cost savings for each type of cargo is interesting; it is explained by the presence of private participation in those ports specializing in bulk cargo.

At the end of 1997, Chile passed a law seeking to modernize state ports and transform the Chilean port system to the new needs of maritime transport. The law intended to introduce more private participation to achieve the objective of modernizing the ports. The law sought to split Emporchi into 10 autonomous public companies to run the 11 state ports, from Arica in the north to Punta Arenas in the south. These new companies act as port authorities, managing infrastructure, but they are not allowed to handle cargo or berthing. The idea is for new port authorities to contract those services with private operators through licenses and concessions.

Even though the actual division of Emporchi did not take place until January 1999, the process of introducing private participation started before then. The main container terminals of the country, located at the ports of Valparaíso, San Antonio, and San Vicente, were concessioned to the private sector in August 1999, for periods of 20 to 30 years.

THE CASE OF ARGENTINA. The reform in Argentina also began with a revision of legislation, with the introduction of significant changes in 1992. Among these changes, decentralization is probably the more remarkable, and Argentina achieved it by closing the central public agency responsible for ports (Administración General Portuaria), and transferring ports to regional governments. It transferred small ports to municipal governments, while splitting the large port of Buenos Aires into three separate zones: Dock Sud, transferred to the province of Buenos Aires; Puerto Sur, which is still to be developed; and Puerto Nuevo, which remains in the hands of the central government.

Other characteristics of the reform process were the deregulation of all port services and the elimination of restrictive working practices. Most important, the reform has introduced significant private participation in building and operating port infrastructure. Given its special role within the maritime transport industry, special attention has been devoted to the development of container terminals, and the reform has attained large improvements in terms of the volume of traffic and productivity.

The port of Buenos Aires has been one of the more innovative world experiences in port reform. The central government-owned infrastructure (Puerto Nuevo) was split into six terminals and concessioned for periods of 18 to 25 years. Initially, the proposed structure was to have five different operators, because a single concession included terminals 1 and 2. As mentioned above, the selection process was designed to avoid a concentration of terminals in the hands of a single operator, forcing bidders to win, at most, only one concession each.

In 1998, market conditions changed, and now only two terminals compete in the container market (terminals 1-2 versus five) within Puerto Nuevo. Terminal 4 is relatively small and dedicated to general cargo, while terminal 3 is multipurpose (general cargo, cars, and passengers). Terminal 6 was forced to close down. Even though the reduction in the number of operators indicates that the process could have been designed better and was probably implemented too quickly, its results in terms of port outcomes have been extremely positive, as can be observed in table 4.9.

The success of seaport modernization in Argentina is in great part due to changes in working practices. In Buenos Aires, private firms have been providing stevedoring services since 1970, but those companies never obtained good results in terms of productivity. Managers were not investing enough in infrastructure and equipment. Plus, seaports unions were strong and kept a separation between stevedoring and loading services, similar to the case of Chile. The Administración General Portuaria was responsible for loading operations, while stevedores had such power that the public agency could not install any new technology that would reduce labor requirements (Raciatti 1998).

THE CASE OF BRAZIL. The situation of the seaport industry in Brazil prior to the reforms introduced in 1993 can be characterized by problems of inefficiency, low productivity, an excess of bureaucracy, and chronic underinvestment. The results of combining all of those problems were port

Table 4.9. Port of Buenos Aires Indicators, 1991 and 1997

Indicator	*1991*	*1997*
Cargo (thousand tons)	4,000	8,500
Containers (thousand TEUs)	300	1,023
Capacity (thousand containers per year)	400	1,300
Cranes	3	13
Operations area (ha)	65	132
Productivity (tons per employee, annual)	800	3,100
Average container time at port (days)	2.5	1.3
Charges per container ($/TEU)	450	120

Source: Estache, Carbajo, and De Rus (1999).

tariffs three to six times higher than international levels, long waiting times for using port facilities, and deficient service provision, which translated into delays in goods' deliverance and reception.

The process of reform started in 1990 with the dismantling of the public agency Portobras and the decentralization of the system. In 1993 Brazil passed a law to establish the general framework of the new reformed port system. This law grants autonomy to all seaports and it allows private participation in cargo handling services, a practice previously prohibited. A movement also started toward the liberalization of port tariffs, with the objective of promoting competition between ports at a regional level.

Reforms have faced strong resistance from port unions, which has been the main factor conditioning and delaying the process. Even though solutions are being implemented to ease that opposition, many ports still have large excesses of workers. Each port now has an institution (Orgao Gestor de Mao-de-Obra) formed of unions and port operators, which is in charge of managing the use and payment of temporary port workers.

Plans call for privatizing 36 state ports, some of which are well advanced, especially for small ports (Itaji, Laguna, Cabedelo, and Porto Velho). The main ports of the country (Santos, Rio Janeiro, and Rio Grande) have been subject to important reforms, and substantial private participation has been introduced through the concession of terminals. At the moment, about 75 percent of infrastructure has been passed to the private sector through concessions. Productivity has increased, ship waiting times have decreased, and the ports have become more competitive. Labor stoppages and other disruptions have been reduced.

In addition, private concessionaires are planning US$1 billion in new investments for the next few years, many of which are contractual obligations and have already been initiated. These investments are mostly destined to modernize or build specialized terminals (coal, minerals, sugar, and so forth) and to buy container handling equipment.

Most of the rest of the public ports in the country are relatively small, because the main export goods pass through privately owned facilities. There are opportunities for the development of new regional hubs for containerized cargo. Two ports could play that role: Rio Grande and Sepetiba. In the near future, both could attract cargo with final destinations in Argentina and Uruguay, and potentially become hub ports for MERCOSUR. Competition for transshipment and final cargo in southeastern Brazil is expected to increase between these new emerging ports and the traditional facilities of Santos and Rio de Janeiro.

SUMMARY OF OTHER INTERNATIONAL PORT EXPERIENCES. This section provides a brief overview of other international port experiences.

Africa

- *Kenya:* Hutchinson Port Holdings signed a contract in 1996 to manage and operate the container terminal at Mombassa. This is considered a first step toward further introduction of private participation at the port.
- *Morocco:* The government's agenda includes seaports' reform, and it is considering allowing private capital into ports. The first stage has been the transformation of the public agency Régie d'Acconage du Port de Casablanca into a new, autonomous public corporation. This change is aimed to improve the efficiency of 11 ports, starting with Casablanca and Mohammedia. A project the World Bank financed has led to a cargo handling productivity increase of 25 percent, and the average dwell time of containers at the port was reduced by 40 percent.
- *Senegal:* The port of Dakar is transforming into a landlord type of organization. Stevedoring services have been transferred to the private sector.

The Americas

- *Bahamas:* The first phase of its modernization project had an objective of providing capacity for container traffic (400,000 TEUs). Now it is trying to start competing against Miami and other Caribbean ports.

- *Colombia:* Seaport laws authorized 25 private ports to handle only specific cargoes. Colombia passed a new general law in 1991, allowing these ports to work on any type of cargo from that date onward. The central public company in charge of state ports, Colpuertos, has started to be dismantled. The objective is to introduce competition among the main ports of the country: Buenaventura, Barranquilla, Cartagena, and Santa Marta. Privatization of the port of Cartagena was initiated after 1991, and since then it has improved its efficiency. The container terminal at that port, Cotecar, has plans for enlargement of its capacity up to 500,000 TEUs/year.
- *Costa Rica:* Private firms perform stevedoring services at the ports of Limón and Moin. At the port of Caldera, however, those services are still under public organization (Alvarado 1998).
- *El Salvador:* The state port of Acajutla is starting its reform process, with plans to introduce private participation.
- *Guatemala:* The ports of Quetzal and Barrios have transferred stevedoring services to the private sector. The process of privatization of those services also has started at the port of Santo Tomás de Castilla (Alvarado 1998).
- *Honduras:* The program for the general reform of the transport sector include plans to privatize seaports. At the moment, Honduras is considering full privatization of all ports, with the exception of Puerto Cortés, which is the main port of the country. The idea is to keep this last port under public control, but to introduce private participation at terminals (Juhel 1994). Private firms have been already authorized, however, to operate stevedoring services at Puerto Cortés (Alvarado 1998).
- *Mexico:* The new 1993 seaports' general law has redefined the role of the state in the industry. The government is relinquishing port administration, terminal operation, and provision of other port services. Privatization started with 26 projects, including the cargo terminals at the ports of Lázaro Cadenas, Manzanillo, Altamira, and Veracruz. The bidding process of seven small ports (Acapulco, Topolobampo, Mazatlan, Puerto Vallarta, Guaymas, Ensenada, and Campeche) followed. The objective is to reach a system of landlord type of ports. Port authorities at each port are planned to manage those publicly owned assets that will be concessioned to them. The decentralization program aims to end up with 22 port authorities running the main ports of the country, with plans to privatize them in the future.
- *Nicaragua:* The public sector performs stevedoring services at the ports of San Juan del Sur and Puerto Cabezas. Authorities at the ports of

Sandino, Bluff, Arlen, Rama, and Corinto are starting to concession those services, however, to private firms organized by port workers (Alvarado 1998).

- *Panama:* This country intends to partly transfer those ports controlled by the state public agency Autoridad Portuaria Nacional to the private sector. Thus, the ports of Balboa and Cristobal were privatized in 1996, and now the Panama Ports Company, a subsidiary of Hutchinson Port Holdings, runs them. Plans also call for an international consortium to invest US$600 million on a project for the construction and operation of a container terminal on the Atlantic side of the Channel at Telfers Island.
- *Uruguay:* The general seaports' law passed in 1992 introduced significant reforms in the ports' activity. The main reform was to make more flexible the use of labor at ports. Uruguay plans to grant a concession of the country's only container terminal to the shipping company Maersk. This proposal is somewhat risky, because the concessionaire should be strictly supervised not to discriminate at the terminal against some rival shipping companies that may compete for import/export traffic. This same company has a concession at the Spanish port of Algeciras for a transshipment terminal, however, and no problems have been reported so far. Both cases differ, though, because other independent alternative terminals exist at Algeciras.
- *Venezuela:* This country initiated a port reform in 1991 by dismantling the public agency Instituto Nacional de Puertos. In 1992, it transferred responsibilities from the agency to the eight port authorities now in charge of ports. These port authorities now have the task of introducing private participation through the concession of container terminals. The new system specifically limits the possibility of any public sector institution (national, regional, or municipal governments) providing services at ports. Caracas and Puerto Cabello were the first ports to introduce private participation (Juhel 1994).

Asia

- *India:* The Indian Ports Association is opening doors to private participation at seaports, though the process is very slow. The only relevant plans are to grant three-year build-operate-transfer contracts to private operators.
- *Philippines:* One of the first concessions in the Asian region was the Manila International Container Terminal, granted in 1988 to a consortium for 25 years, and involving investments for $54 million. Plans call for

enlarging further the capacity of Manila up to 1 million TEUs/year, with a fifth berth. Regarding other ports, Asian Terminals Inc., a joint venture formed by P&O and a local firm, has a 15-year contract to operate the South Harbor of Manila; Hutchinson Port Holdings won a bid to develop and operate a container terminal at Subic Bay, but the deal was thereafter rescinded among political turmoil over lack of transparency.

- *China:* Shanghai's port authority and the private firm Hutchinson Whampoa Ltd. formed in 1993 a joint venture to create the company Shanghai Container Terminals Ltd., which owns and operates the container terminals at the port. During its first year in operation, the company handled 25 percent more containers than in the previous year, and productivity increased more than 30 percent. Investment plans of $673 million are being completed. Hutchinson Whampoa is also involved in the development of the Yantian port, a location intended to become one of the four China hubs, together with Dalian's Daiyo Bay, Beilun, and a port in the Fujian province. In Dalian, PSA Corporation has formed a joint venture with the port authority to operate the container terminal. The parties that are more interested in developing Chinese ports are shipping companies with regular services in the region and private investors based in Hong Kong (Peters 1995). For example, Maersk has a preferential agreement as the main user at the port of Yantian, and P&O has made significant investments in the container terminal at Shekou and in the development of the Tianjin port.
- *Malaysia:* This country initiated privatization in 1986 at the port of Kelang, when the container operations were granted for 21 years to the joint venture Kelang Container Terminal (KCT), formed by the port authority and private investors. In 1992, a second privatization phase launched, concessioning the rest of the port's infrastructure to another private company (Kelang Port Management, or KPM). Although the port authority does not have participation at KPM, it holds a golden share to keep control over essential issues. The KPM container terminal thus competes with the KCT terminal within the port of Kelang. A project is developing new facilities at Westport, located also in Kelang. The government designed the initial plan to reduce congestion at existent facilities, and it intended to finance it with public funds. But eventually it was concessioned to a private consortium (Kelang Multi-Terminal Consortium) with a 30-year term contract. These new facilities that the consortium is developing will compete against those of KCT and KPM, reinforcing even more the competitive framework of the Kelang port.

Price Regulation

The economic regulation of ports aims to put some limits to the market power that private port operators may enjoy by having exclusive rights. This type of regulation basically is performed by establishing some limits on the tariffs that are charged to port users, or on the total revenues that private firms obtain. Because the types of tariffs involved in the use of ports are diverse, analyzing separately the regulation over port dues, over charges for cargo handling, and over fees that concessionaires pay for the use of infrastructure, is convenient.

Port Tariffs

In seaport activity, the users of a port must pay diverse charges for the services they receive and for the use of facilities. Port tariffs (or port dues) are charges on ships for the use of the general infrastructure of a port. Port authorities impose these tariffs, although they do not always directly collect revenues, a task that in some cases concessionaires perform. Apart from infrastructure, port dues may include charges for the use of compulsory berthing services (pilotage, towing), particularly at ports where the port authority is in charge of those services. Another part of the total fee stems from tariffs on all cargo that passes through the port's facilities. Shipping companies partly pay these tariffs, and the rest is directly charged to shippers.

Although port tariffs are relevant when shipping companies and exporters/importers choose between ports, their weight is relatively small compared to the total cost that port users must bear. Table 4.10 shows that the largest part of the bill is cargo handling (loading/unloading, stevedoring, storage, and so forth).

Table 4.10. *Relative Weights of Different Port Charges*

Type of charges	*Percentage of total bill*
Port tariffs on the use of infrastructure	5–15
Berthing services	2–5
Cargo handling	70–90
Consignees	3–6

Source: Suykens (1996).

Port industry experts are generally of the opinion that the elasticity of the demand for port services with respect to port tariffs is relatively small (Slack 1985). For shipping companies, relevant factors when choosing a port are the general quality (equipment, waiting times, operating times, and so forth.) and the existence of business opportunities (demand for cargo transport from exporters and importers). For the shipper, the important variables would be the charges on cargo handling, the frequency of regular services (liners), and the existence of charter services from the port for special shipments.

As a consequence of these behaviors, one can conclude that port authorities can raise and lower port tariffs within a wide margin, without affecting their demand levels. An exception would be a region with fierce competition among ports with similar facilities and inland connections. In that case, it is possible that a slight variation in port tariffs could lead to traffic deviations, and thus render port tariffs as a strategic variable for competition.

In theory, the optimal rule for determining port tariffs for the use of port infrastructure is to make users pay the marginal costs that they generate. As with other industries in which infrastructure is publicly used (such as airports and roads), however, marginal costs are very small, because they only consist of maintenance and repair expenses. These costs are small when compared to construction costs. Therefore, cost recovery is a problem if employing the social optimal pricing rule.

A classical solution to this problem is the public sector paying for infrastructure costs, and thereafter, the users only being charged the marginal cost. The argument for the use of this system is that whether users pay full costs, some of them could be driven out of the market, even if they are prepared to pay the marginal costs. But because the option of public sector help for seaports is not available in most countries, new solutions to determine port charges are needed. One possible alternative is to use the concept of long-run marginal cost, which keeps the idea of social optimality, and at the same time, achieves full cost recovery (see box 4.2).

In practice, port tariffs are determined by rule of thumb and do not necessarily relate to investment costs or opportunity costs for the use of infrastructure. As a general rule, port authorities seek real return rates on assets of around 8 to 12 percent during the economic life of the infrastructure (DSC 1998).

Tariffs charged on ships for the use of infrastructure usually depend on some capacity measure, such as gross registered tons or some other alternative. At some ports with high demand levels, one can establish port tariffs on ships on other criteria that better reflect the opportunity cost for the port authority of having a given ship use a piece of infrastructure (for example, at the port of Rotterdam, ship tariffs partly depend on their total

Box 4.2. Concept of Long-Run Marginal Cost

The concept of long-run marginal cost (LRMC) has been used in sectors with expensive infrastructure assets, such as roads or public utilities (water, gas, and so forth). The problem all these industries share is that significant investments are needed to build infrastructures, which constitute large sunk costs, while marginal costs of operation are relatively small. If the socially optimal pricing rule is applied—price equal to marginal cost—recovering investment costs is not possible. A second-best alternative to set low prices, and to avoid excluding potential users, is to use LRMC as a basis for pricing. LRMC is defined as the sum of short-run marginal cost (SRMC) plus marginal cost of capacity (MCC):

$$\text{LRMC} = \text{SRMC} + \text{MCC}$$

For the case of seaports, the marginal cost of capacity would be the additional cost of infrastructure required to service one more unit of cargo above the maximum port's capacity. For ports with excess capacity, each additional user does not require new infrastructure, so in that case, MCC = 0, and long- and short-run marginal costs would be equal. Meanwhile, for the case of a port with congestion problems, the marginal cost of capacity is positive, and therefore, LRMC > SRMC. For ports, SRMC is formed only of maintenance and infrastructure repairing costs.

Some difficulties in practice exist for LRMC pricing: (a) infrastructure cannot be continuously enlarged (indivisibilities are derived from berths' minimum sizes), and (b) infrastructure assets have long economic lives. If the rule of setting price equal to long-run marginal cost is applied, port tariffs could oscillate dramatically between years, because users calling at a port in periods of capacity enlargement would then be paying for assets that are thereafter used during long periods. In practice, a solution is to use some formula to distribute the cost of construction, plus its associated financing cost, during the economic life of the asset. Thus, what part of total cost of capacity port users should pay each year is estimated, so that port tariffs do not vary too much, and at the end of the period the users have financed the infrastructure's construction.

length). Imposing extra charges for ships with special requirements is also possible, in terms of draught or other characteristics.

For tariffs on cargo, port authorities usually discriminate among types of goods, in some cases following complex classification schemes (in European ports, one can find countries with lists of up to 56 different types of cargo and charges). The origin of these systems of charges is the idea of extracting as much rent from users as possible (the traditional practice of charging what the market can bear). Therefore, prices are more or less related to the value of the goods passing through the port. A trend, however, at least in Europe, is toward reducing tariffs on cargo to attract shippers, and increasing tariffs on ships to balance port authorities' total incomes.

In terms of regulatory needs, because most world ports are going to a landlord type of model and not to full privatization, imposing regulation, in principle, is not needed on port prices. Nevertheless, in the case of ports without competition in their region, a regulator should possibly supervise the charges on the use of infrastructure, because in such cases the ports enjoy a monopoly position. The need for regulation arises even if the port authority is a public institution, because a risk of capture by third parties is still possible, which can lead to a nonoptimal tariff set by the port authority.

Meanwhile, regulating charges on port users for complementary services (such as berthing) is needed in cases in which private operators provide these without guaranteed competition. A system based on maximum prices is usually employed to regulate this type of service, because no significant problems occur in estimating their costs.

Cargo Handling Charges

As mentioned earlier, charges for cargo handling services are the most relevant for port users. Therefore, because these charges affect ports' competitive position, relating these charges closely to the real costs of service provision is crucial. The whole process of privatization and liberalization is aimed at making these charges to be in accordance with market mechanisms, instead of set by public institutions, as traditionally was the case in most ports worldwide. The liberalization process does not guarantee that market mechanisms will prevail, however, so regulation is still needed in ports where competition is absent.

In general, large ports tend to liberalize their cargo handling charges so that private operators can freely set them. Regulation is not required in those cases, because operators must adapt their charges to market conditions. Port authorities usually keep some form of control, however, such as setting a maximum level of charges. The need for regulation is greater in the cases of medium and small ports, because the size of the market only allows for a limited number of operators, a situation that can easily result in collusive practices among them.

One should study in detail the maximum charges authorized to port operators by port authorities or the regulatory institutions, because this is the basic instrument of regulation over private operators. In practice, concession contracts that port authorities and private firms sign are usually not too precise on determining those authorized maximum charges. Generally, contracts state that the private operator will be allowed to obtain an adequate rate of return (as in the concession contracts in Buenos Aires), but they do not specify how this principle will be implemented in practice.

Concession contracts should explicitly include rules that a regulator is going to follow when setting maximum charges. Because one knows that port operators are going to adapt to the type of regulation, and the effects of price caps and rate of return regulatory systems are not the same, the regulator must choose a mechanism according to both the information available and its own objectives. Thus, for example, if port operators are intended to make substantial investments in equipment, establishing a rate of return type of regulation is adequate, because that will provide incentives for them to invest in capital assets.

Difficulties in evaluating the costs of concessionaires operating at ports can, in principle, be overcome by establishing some form of regulation based on comparisons between different operators (yardstick competition). This implies analyzing cargo handling charges among similar seaports in a given region, and trying to derive conclusions in terms of cost efficiency and charging practices. Some established international reference values already exist for some types of cargo, which regulators can use to get an idea of the outcomes that should be expected from a private operator. Nevertheless, these reference values must be adapted according to local conditions (average wage levels, interest rates, and so forth). Charges on cargoes may vary greatly across ports, according to the type of technology and the age of the equipment employed for cargo handling.

For example, with containers, the price charged per TEU is an easy variable to obtain, and a regulator can use it after some adjustments for local conditions. The reference values can serve as a benchmark for the efficiency that a private operator should be able to achieve, which then can be used as a limit for imposing price cap regulation (see table 4.11). In the case of the container terminal at the port of Santos (Brazil), regulation has been imposed on the new private concessionaire in the form of a target for the charge per TEU. In a period of three years, the price must be lowered from US$550 per box to US$150.

Concession Fees

A relevant question when introducing private participation in seaports is to determine the payments that the operators must make to the port authority or the agent that owns the infrastructure assets (named concession fees or canons). Even if those fees do not directly affect port users, evidently the higher the payments private operators make for the use of infrastructure, the more income port authorities receive, and port tariffs can

Table 4.11. Comparison of Container Handling Charges Across World Regions, 1996

Region	*Port*	*Price per loaded TEU (US$)*
North Europe	Antwerp	120
	Felixstowe	173
	Hamburg	182
	Rotterdam	156
	Zeebrugge	123
South Europe	Algeciras	193
	Barcelona	211
	La Spezia	240
	Marseilles	233
	Pireus	203
Asia	Pusan	175
	Kaohsiung	140
	Manila	118
	Singapore	117
North America	Halifax	168
	Los Angeles	256
Australia	Melbourne	199

Source: DSC (1998).

then be reduced accordingly. Private operators will then try to pass their higher costs to users through their cargo handling charges, however, so port authorities should establish a careful balance regarding those prices that they can directly control.

An advantage of this mixed form of revenue for port authorities (from port tariffs and concession fees) is that part of the demand risk is left to the private operators, which then have correct incentives to provide efficient, low-priced services to minimize that risk. In addition, concession fees provide port authorities with a safe, continuous cash flow, therefore providing the possibility to finance general port costs or even part of the facilities' construction/rehabilitation costs.

In European ports, revenues obtained from port tariffs are in general higher than revenues from concession fees, with these latter averaging 37 percent of total income received by port authorities. Examples exist in other regions, however, as in the case of the port of Baltimore (U.S.), where more than half of total income is obtained from concession fees (58 percent), therefore indicating a higher presence of private operators (PDE 1998).

No established procedures determine the level of concession fees for private firms to pay. An optimal rule should be to relate payments to the opportunity costs of the infrastructure and those superstructure elements with which the concession might be associated. For infrastructure, an approximation for the opportunity cost could be the market price of the port's adjacent land, modified by the specific characteristics of the surface used by the concessionaire. Meanwhile, the opportunity costs of equipment granted by a concession are easier to estimate, because they are equal to their price in a rental market.

One can add other aims to the basic objective of concession, for example, sharing the risk of demand fluctuations between operator and port authority. This risk could be shared by making concessionaires' payments dependent on their level of activity, with some minimum guaranteed payment. The optimal system to determine concession fees is a mixed combination of opportunity costs and risk allocation objectives.

In practice, however, port authorities do not often use any market criteria to determine the opportunity cost of assets. Concession fees are usually fixed payments per square meter used, which are periodically revised (see table 4.12). Fees usually depend on the service the concessionaire provides, so that the price per square meter is different if the surface is used for container handling than if it is devoted to specialized storage areas. In some cases, fees depend on the volume of demand attended to by the operator, therefore achieving the objective of risk allocation. For example, in the case of the concessions of the port of Buenos Aires, concessionaires pay according to the total volume of cargo handled, with a guaranteed minimum payment for the port authority.

Quality and Safety Regulation

As important as the economic regulation, the public sector must keep an eye on quality and safety issues at privatized ports. This is an important task, because the activities performed at ports are prone to cause externalities to the environment and to other port users. Private concessionaires are not likely to properly consider these externalities unless they are forced to by some established, clear rules.

Congestion Problems

The waiting time of ships is one of the port characteristics that shipping companies value when choosing between ports. The total time that a ship stays at a port is equal to the sum of the time employed to obtain

Table 4.12. Concession Fees for Different Ports, 1997

Port	Annual price per sq.m. (US$)	Revision frequency	Revision mode
Baltimore	6.5	Annual	Variable
Bremerhaven	2.3	5 years	Price index
Bordeaux	4.5	Annual	Price index
La Spezia	5.7	Variable	Variable
Le Havre	3.8	Annual	Price index
Lisbon	15.0	Annual[a]	Price index
Oslo	61.5	Variable	Variable
Rotterdam	3.2	Variable	Variable

a. It also charges a variable fee on volume: US$0.3 per ton or US$5.5 per container.
Source: PDE (1998).

the required services and supplies, and this time must be considered as a cost for port users. The generalized cost paid by a ship using a port is equal to monetary charges (port tariffs + cargo handling charges + other services' prices) plus the time spent in the provision of services (see box 4.3). Therefore, the shorter the waiting time, the lower the generalized cost of port use, and the more attractive the port is to users.

The first part of total ship waiting time is the time spent at the port maritime access zone waiting for a berth to be available. Two scenarios are possible. First, the port could suffer from no congestion problems; therefore, the waiting time is equal to zero. The second situation occurs in ports with congestion problems, which can result in relatively long waiting periods (up to several days).

In the latter situation, the long-term solution is to enlarge the port's capacity, but in the short term, using port tariffs can establish some demand rationing. Raising these tariffs could induce some users to seek alternative facilities, therefore decreasing average waiting times and improving the welfare of the remaining users. During the construction period of new infrastructures, however, the port authority should try to minimize the disruption for port users, because otherwise some traffic segments can permanently deviate to rival ports. If one does not correctly manage congestion periods, investments in new facilities could only result in excess capacity.

For ports that are contracted out to private management, or that are fully privatized, a regulator should be concerned with guaranteeing that decisions

Box 4.3. Concept of Generalized Cost

As in other transport modes, when analyzing the cost that users incur, we have to consider not only the monetary cost of the fare or tariff, but also the value of time spent to obtain the desired service. In the case of seaports, ships are charged by different concepts (such as use of infrastructure, berthing, cargo handling services, and other supplies), but they also spend considerable lapses of time at ports waiting to be served. In a broader definition, we should also include such things as costs suffered from cargo damages or losses, but by considering only prices and time, we can define the generalized cost for port users as follows:

$$\begin{aligned}\text{Generalized cost} &= \text{price} + \text{time} \times \text{value of time}\\ &= (\text{port tariffs} + \text{services' charges}) + t_{\text{ship}} \times V_{t\,\text{ship}}\end{aligned}$$

In the expression above, t_{ship} would be the total time spent by the ship in obtaining port services, from when it enters the port until it exits, and $V_{t\,\text{ship}}$ would be the opportunity cost of the ship per unit of time (rent that is lost when the ship is not providing transport services).

on port capacity are made sensibly. A private manager, in principle, will not have a long-term perspective on running the port if the management contract has a fixed term and no renovation is expected. Thus, if the performance of the port is measured by its financial results, a private manager could leave congestion problems unresolved, and simply obtain extra income from high port tariffs (which some users are prepared to pay in the case of congestion).

Quality of Service

The second relevant component of cargo ships' waiting time, and therefore of their generalized costs, is the time spent being loaded/unloaded, which an efficient port should try to minimize. Moreover, safety procedures also should be followed to avoid damage to cargo. Therefore, regulation on private concessionaires should not only be concerned with prices, but also include quality of service provisions in the contracts.

In principle, a private operator would be interested in cargo handling services being provided quickly and safely, for its clients to be satisfied. However, in some cases, a profit-oriented operator may not care excessively about safety and only value speed (at terminals with high demand), or spend too much time servicing ships with expensive cargo that are prepared to pay high charges, which raises costs for other clients with low-value cargo waiting to be serviced (cream-skimming problem).

To provide incentives for loading and unloading services to be done as efficiently as possible, and to avoid situations such as the two examples above, the concession contract can include minimum standards on safety and servicing times. For example, including a variable part on the concession fee is possible, which could depend on ships' average waiting times. By using this instrument, the private operator would have incentives to service ships optimally, and to invest in the required equipment to reduce those waiting times as much as possible. Similarly, one could impose penalties if a safety standard indicator falls below a certain minimum (for example, amount of cargo damaged or lost).

The quality of cargo handling services also involves some technical aspects such as spending adequate amounts on the maintenance and repair of equipment. Periodic revisions must be performed on the equipment to guarantee the minimization of accidents and disruptions. Because these revisions are costly but can improve the general efficiency of the port, the concession contract should explicitly include some conditions. Other safety aspects that a well-designed contract must include are obligations for the concessionaire to maintain sufficient lighting in the terminals for night services, adequate ramps for passenger services, and separate facilities for cargo and passenger services.

Safety

A high density of vessel traffic in the access zones of a port and within its area increases the risks of collision and ship stranding, especially in stormy conditions. Given the negative externalities that maritime accidents cause on other port users, and the potential environmental consequences, regulation on general port safety and quality of services related to ships' movements must be strict, and compliance closely monitored.

All ports generally make pilotage use compulsory for vessels above a certain capacity or dimension and for ships transporting dangerous cargo. When pilotage is imposed, a technical expert with a knowledge of port characteristics (a pilot) should be on the ship as it enters and exits, or at least the captain must follow instructions by radio.

When the port authority does not directly provide pilotage services, but independent agents offer them, the port regulator must somehow control this activity. First, some economic regulation must be made on the tariffs that pilots charge for their services to shipping companies, particularly when sufficient competition among several agents is not guaranteed. Second, technical capacity must be ensured by requiring pilots to demonstrate their ability to perform the required tasks. One can use a system of licenses to regulate these safety aspects. In addition, setting minimum equipment standards (such as boats and radios) is recommended.

Pilots are the agents who determine the number and power of tug boats that a vessel requires to perform movements to enter and exit the port. Therefore, collusion between pilots and towing firms is a risk, and users may be forced to buy extra services that are not necessary. To prevent this, ports must have clear regulations on the minimum requirements for towage services available to all port users (De Rus and others 1995).

In ports where private firms provide berthing services (safe tying of vessels to berths), port authorities should provide regulations to guarantee that safe procedures are correctly followed. Incidents have occurred in which insufficiently tied ships have drifted within the port area, causing accidents. This problem is especially serious for tankers, because sudden unberthings when delivering or receiving supplies can lead to dangerous spills.

Finally, port authorities must always have emergency plans in the event of accidents, and port workers must be trained on evacuation procedures. Concessionaires should be obliged by their contracts to fulfill minimum safety requirements in their buildings and superstructure elements: emergency exits, fire-fighting devices, signs, and so forth.

Performance Indicators

To evaluate the outcomes for a seaport, several types of indicators can be used. These indexes are useful if they can be easily computed with available information (from port authorities and concessionaires), they can be updated regularly to study the evolution of the port over time, and they have some regional benchmarks against which they can be contrasted. Using these indicators, a regulator can assess the performance of a port, and evaluate if the results that concessionaires achieve are satisfactory.

Because one can compute many indexes from ports' information, classifying the possible indicators into three separate groups is useful, according to the aspects that they aim to measure: (a) physical, (b) factor productivity analysis, and (c) economic and financial.

Physical Indicators

The type of information that this set of indicators tries to measure is conceptually simple. The idea is to measure how much cargo is moved by a port, how fast ships are serviced, and how quickly cargo is transferred to other transport modes. Therefore, the basic indicators are time measures and, indirectly, the total volume of traffic that the port receives. The most commonly used physical indicators in the seaport industry are as follows:

- *Ship turnaround time*: This is the total time that a vessel spends at a port, from entrance to exit. One can divide this turnaround time into two parts: time at berth and time outside. If a port does not have this detailed information for all vessels, computing some average turnaround time is always possible by dividing some estimated total vessel stays over the number of vessels calling at the port during a particular period.
- *Waiting rate*: Using the two types of times described above, one can figure the waiting rate as the time in the port outside the berth divided by the time at berth. This index provides information about congestion problems at the port. A high value indicates that ships must spend a significant part of their port time waiting for a berth space to be available.
- *Berth occupancy rate*: This represents the percentage of total available time that berths are in use by ships. This is a useful indicator for obtaining an estimate of the level of a port's activity. It must be complemented with additional information, however, such as the turnaround time, because a high value for the berth occupancy rate is a positive indicator (showing that a port is busy most of the time), but only if the turnaround time is low. Otherwise, this could be regarded as an extremely inefficient port, whose users spend too much time berthed but not serviced.
- *Working time over time at berth*: This is another indicator that complements those above. A value close to 1 indicates that a ship is being serviced for most of the time that it spends in port. A smaller value reveals that the ship is idle most of the time that it is berthed (with the corresponding opportunity cost). If detailed information is available, knowing the distribution of the remaining time (time at berth minus working time) is also interesting. Some ports have records on the idleness due to rain, strikes, equipment failure, and other reasons.

Because most factors affecting this list of indicators depend on the type of ship and cargo, providing benchmark values valid for every ship and port is difficult. A solution to have valid reference values is to compute them separately by vessel type: bulk carriers, containers, and general cargo. For example, for the waiting rate, the best values observed in the world are 5 percent for container ships and 20 percent for bulk carriers. These are obtained in large ports that operate as regional hub centers (Rotterdam, Antwerp, Felixstowe, and Singapore).

Another interesting performance indicator, from the viewpoint of shippers that export/import goods, is the time required for cargo to pass through the port:

- *Cargo dwell time:* This is the time elapsed from when cargo is unloaded from a ship until it exits the port, or vice versa. It is usually measured in days, and naturally, the smaller the value of the index, the higher the port's efficiency.

A high value for this indicator reveals cargo management problems, and although knowing the cause of a long stay for shipments at a port would be extremely interesting, having information in that much detail is usually difficult. The best practices are generally obtained in the container market, where large ports exhibit values around 4.7 days. Meanwhile, the dwell times are longer for general cargo, averaging 7 to 12 days.

Causes for delay can be due to the poor performance of administrative services, such as customs or sanitary inspections, or they could originate through poor coordination between ship and land modes of transport. The presence of delays that increase the cargo dwell time can be disastrous for some kinds of goods, such as fruits, vegetables, and fish.

Finally, other types of indicators that the physical group could include are those related to safety concerns, such as the number of accidents or incidents suffered by ships at a port. In order to be accurate, these indicators preferably should be expressed relative to an exposure-to-risk variable, such as the total number of ship movements to and from the port. To evaluate the safety commitment of concessionaires, computing the investments on safety over total expenditures, or over a volume of cargo handling, is recommendable.

Factor Productivity Indicators

In addition to physical indicators that provide information on ports' efficiency, having some knowledge of labor and capital productivity is important, so that when one detects low efficiency, identifying the reasons causing it is possible. Some simple indicators to measure productivity are as follows:

- *Tons per worker-hour or per gang-hour:* These measures are aimed at measuring labor productivity, but when making comparisons across ports, one must be ensure that conditions are similar, because, for example, the size of a gang can vary between two ports. Similarly, when comparing worker productivity, one should do this only for equivalent types of cargo. Moreover, the information would have to be complemented with indexes for the state and type of equipment employed, because labor productivity varies according to a port's capital stock.

- *Tons per crane-hour:* This simple indicator evaluates the productivity of the main equipment for cargo loading and unloading. In order to make comparisons across ports, one should guarantee some homogeneity on the type of cranes. For containers, comparing ports is easier, because both cranes and cargo are basically homogeneous. For this type of cargo, using TEUs as the unit of reference is preferable.
- *Tons per linear meter per year:* This indicator provides a measure of a port's efficiency in using its basic infrastructure to provide services to ships.
- *Tons (or TEUs) per ship-hour:* This indicator gives an idea of the total productivity of a port in cargo handling. A reduced value for the index will indicate low efficiency and the imposition of longer times on ships.

Economic and Financial Indicators

Lastly, a third group of indicators can be calculated to provide regulatory institutions with a complete picture of a port's situation. The objective of all of these indexes is to reflect port finances and level of charges to users:

- *Operating surplus over gross registered tons/net registered tons* or *operating surplus over handled ton.*
- *Total income (expenditure) over gross registered tons/net registered tons.*
- *Charge per TEU.* An index to evaluate the efficiency of a port in handling containers is the total charge per TEU. This is becoming an international reference benchmark, though one recognizes that local conditions over some particular costs (such as labor) may vary considerably. Therefore, using this indicator on a regional basis is recommended. Overall, best practices worldwide indicate that the minimum for this index can be between US $120 and $150 (see table 4.11).

Conclusions

Many international experiences have shown that the effects derived from the introduction of private sector participation in ports are highly positive.

Improvements at the operational level are obtained (better and faster services, reduced waiting times, and so forth), in addition to significant increases in investments that are reported after a port is reformed and its activities are transferred to the private sector. In fiscal terms, many port systems worldwide that once drained resources from public budgets have changed. They now maintain sound finances, which in most cases allows them to contribute positive revenues to the treasury.

Options for the participation of private firms at ports are diverse, but probably the most typical today is the use of concessions for port terminals (either through leases or build-operate-transfer contracts). Because the number of terminals within a port must be necessarily reduced, the challenge for the new port systems is to promote an adequate competitive environment for private concessionaires to provide port services efficiently and to set tariffs according to real costs. In some cases, the size of the port will allow competition among operators within the port. Regulators have important tasks, however, for small and medium-size ports. The first is to promote competition for the market by designing an optimal auction to award contracts. Regulators must also study if competitive conditions exist at a regional level (users may have the option of alternative terminals at other ports if a private operator imposes high tariffs). If that is the case, the task of port regulation will be much easier than in situations in which competition is absent.

Concessions for port terminals are, by definition, long-term contracts, so correctly designing and monitoring all the details required for a successful partnership between the public and private sectors is important. This chapter has reviewed all the necessary elements to be included in a concession contract for a regulator to promote the maximum degree of competition and to choose the best options according to the objectives pursued. Some indicators of port activity are provided, which in the future should be improved and incorporated as a standard toolkit for regulators. A database for these indicators, calculated for different ports around the world, would be extremely useful for the work of port regulators, who could use the indicators as reference benchmarks to impose targets on private concessionaires.

5

Railways

Javier Campos and Pedro Cantos

With the rail industry transformed worldwide, regulation of the sector should remain simple and flexible to protect its share of transportation markets. Apart from providing a stable legal and institutional framework and fostering competition and market mechanisms, regulators should refrain from intervening in the market—unless the goal of economic efficiency (subject to the socially demanded level of equity) is in jeopardy. This chapter reviews these ideas.

Characteristics of Railway Services

The rail industry poses a number of specific problems for transport economists and regulators that are only partially shared with other transport modes. These elements are the multiproduct nature of the activity, the particular cost structure of railroad companies, the role of infrastructure and networks, the existence of indivisibilities in inputs and outputs, the organization of rail transport as a public service, and the existence of externalities in the transport system as a whole. According to Button (1993), these characteristics define a descriptive framework for this sector, and they jointly determine the main factors that one should consider when studying in detail the appropriate economic regulation for the rail industry.

The Multiproduct Nature of the Activity

Rail companies are, in most cases, multiproduct firms that provide different types of freight and passenger transport services. In the case of freight, along with the usual transport of bulk freight, rail operators also supply complete cargo wagons or trains, parcel and postal services, and other services of intermodal transport. In the case of passenger transport, long-distance traffic usually coexists with local services (suburban and commuter trains), regional services, and in certain cases, even with high-speed trains.[1]

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

At the cost level, another important aspect to consider in the multiproduct setup of the rail industry is the sub-additivity of the cost function faced by a railroad. According to Baumol (1977), a cost function is subadditive when the provision of services by a single firm is more efficient (in terms of a lower unit cost) than the same production carried out by two or more companies. This idea conveys two relevant implications for the rail industry. First, is it more efficient for a single firm, rather than two separate firms, to supply both infrastructure and transport services? Second, if the infrastructure and services are separated, is the supply of such services more efficient within the context of a monopoly, or should two or more firms participate? This analysis, connected to the advantages and disadvantages of the separation of infrastructure from services, will be discussed in depth.

Railway Costs

Waters (1985) broadly distinguishes four railway cost categories: (a) train working costs, including the cost of providing transport services (fuel, crew, maintenance, and depreciation of rolling stock); (b) track and signaling costs

---

1. This chapter will not analyze commuter and suburban passenger traffic, because they should be studied within the more general framework of urban transport.

(including operation, maintenance, and depreciation of infrastructures); (c) terminal and station costs; and finally, (d) administration costs.

The first two categories are prevalent in most companies and change according to several factors.[2] Among train working costs, for example, rolling stock costs depend on both their number and the distance they run. Fuel costs depend on car-kilometers run for each type of vehicle, while train crew costs vary according to train-kilometers run. Track and signaling costs usually rely on the length of the route (because they typically request a single, standard-quality track). The amount of track and signaling needed, however, changes with the number of trains requiring paths, although this relationship is not constant. Terminal and station costs depend on traffic volumes, but they vary considerably with the type of traffic. For instance, bulk freight handling requires more terminal expenses than parcel services. Similarly, long-distance passengers require more services (ticketing, reservations, luggage, and so forth) than short-distance users. Administration costs fluctuate depending on the overall size of the firm, although the precise nature of this dependence is generally difficult to determine.

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

Rail Infrastructure

Since the birth of the rail industry in the 19th century, mainstream economists have always considered that the larger the size of a railway company, the greater its efficiency. The existence of substantial fixed costs (particularly those associated with infrastructure) traditionally led economists to assume the presence of important economies of scale, and thus to regard rail transport service as a textbook example of a natural monopoly.

This notion has been heavily challenged in recent decades, however, by the introduction of new ideas into the industry's economic analysis.

---

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

Particularly, the upheaval of the theory of contestable markets (Baumol, Panzar, and Willig 1982) contributed to clarifying the proper definition of the natural monopoly concept in terms of the sub-additive cost function. This concept implies that duplicating rail infrastructure is generally inefficient (and is therefore subject to natural monopoly conditions), but once the network has been deployed, more than one company can efficiently cover the cost of operating rail transport services and rolling stock, either as actual or potential competitors.

Therefore, from the regulatory point of view, the conclusion is that one can deal with infrastructure and services in different ways: the former, as a natural monopoly (at least, when the infrastructure has not yet been built, although not necessarily after that moment), but also as a potential provider of adequate access to any willing-to-serve operator; the latter, as any other competitive economic activity that could be provided by multiple competing operators or by a single firm under some sort of concession or license arrangement.

Asset Indivisibilities

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

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

A final implication of the indivisibilities in the rail industry's capital assets is that innovation and infrastructure improvement projects are usually deferred and only carried out in small, discrete amounts. Railway firms seldom change the entire definition of their existing network, which in most countries corresponds to an inherited burden from past decades when the traffic structure was very different from today. Instead, they opt for partial renovations that often introduce technical asymmetries between tracks within a country or region, and accentuate indivisibilities and inflexibilities (Boyer 1997).

Railway Transport as a Public Service

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

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

Externalities in the Transport System

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

The current intermodal misallocation (more road users than rail users) arises from the fact that road transport does not fully internalize all of the social costs that it generates. Economists often recommend the use of congestion and/or pollution rates to account for this. When these mechanisms are not feasible or politically viable, however, decreasing railway fares to improve the overall intermodal balance might be preferable, which is an additional consideration for rail regulation.

In summary, all the foregoing characteristics, shown in table 5.1, suggest that regulation of railway transport should be analyzed within a general context, taking into account the industry's technological and organizational features, beginning with a detailed evaluation of recent performance.

Table 5.1. *A Summary of the Economic Characteristics of the Rail Industry*

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

Source: Authors.

Privatization and Regulatory Trends

Table 5.2 summarizes the overall evolution of rail transportation in recent years as compared with other transport modes for Organisation for Economic Co-operation and Development (OECD) countries. The 1970s and 1980s saw a substantial fall of market share in both freight and passenger markets, which stabilized during the 1990s. The decline is particularly relevant because it was during a period when the total volume in both markets grew about 50 percent, implying that the rail industry was not able to take advantage of growing demand in the past 25 years.

This substantial reduction in market share is not specific to OECD countries, but is a common trend worldwide. It can be attributed to both exogenous and endogenous causes. The former include the rapid development of alternative modes of transport, especially road. For passengers, economic growth fostered the development of the automobile market, leading to enormous growth in motorization. In freight transport, the expanding, competitive trucking sector gained a growing percentage of transport in many countries. For example, in 1970 in Europe, the number of cars per 1,000 inhabitants was 150, a figure that now is 424. Similarly, the number of heavy vehicles and trucks increased from 7 million to 17 million from 1970 to 1994.

The endogenous causes of the decline can be summarized in the inability of the sector to adapt to the changing conditions of the economic environment. Regulation remained obsolete and the rail industry was slow to react. The policies adopted during the 1980s did not halt the steady loss of

Table 5.2. *Market Shares of Different Transport Modes, Selected Years, 1970–94*
(percent)

Type of transport	*1970*	*1980*	*1985*	*1991*	*1994*
Passenger					
Rail	10.4	8.6	7.3	6.9	6.9
Private car	77.3	80.0	83.4	84.4	84.4
Bus	12.3	11.4	9.3	8.7	8.8
Freight					
Rail	31.3	23.2	21.2	17.9	15.5
Road	55.2	65.9	69.3	74.0	76.2
Waterways	13.5	10.9	9.5	8.1	7.9

Source: CEMT (1996).

market share, the growing financial deficits, and, in some countries, the impossibility of raising the low productivity indexes of the industry. Thus, more radical restructuring processes were put into practice.

The Traditional Model and Regulation of the Industry

During the past 50 years, the most common market structure in many countries' rail sectors was a single, state-owned firm, entrusted with the unified management of both infrastructure and services. Despite some differences in their degree of commercial autonomy, the traditional methods of regulation and control of this sort of company have been relatively homogeneous. In general, it was assumed that the monopoly power of the national company required price and service regulation to protect the general interest. In addition, the companies were obligated to meet any demand at those prices. The closure of existing lines or the opening of new services required government approval. Thus, competition was rare and often discouraged, and preservation of the national character of the industry was considered the key factor governing the overall regulatory system.

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

These distortions did not come from any artificial reduction in the range of services provided, or from excessively high fares, but more commonly from an unjustified increase in the supply of services (and hence, of costs). Such behavior implied larger public subsidies. In many cases, the lack of commercially oriented tariffs and investment policies explained many of the difficulties faced. Together with the burden imposed by the technical

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

characteristics of the sector, this placed most railways in a weak position to compete against alternative transport modes. Fierce intermodal competition, however, was not able to improve the competitiveness of the railway system by itself. Adopting measures affecting the internal behavior and structure of the sector itself was necessary. Therefore, the sector's overall decline sparked a widespread restructuring movement around the world.

The Movement toward Privatization

The worldwide restructuring process of the rail industry began with timid reforms. For example, many countries began by replacing their national railways with autonomous commercial bodies possessing independent, realistic balance sheets, in which the government could explicitly subsidize public service obligations. Other countries opted to substitute their old geographically-based management with a multidivisional structure, defined by the companies' different lines of business or services. Table 5.3 allows us to compare similarities and differences among several countries.

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

The changes have involved revising laws and other regulations affecting railways, reducing staff, dealing with pension issues, and deciding how much property the state should sell and how much it should retain. In addition, several arrangements for paying for unprofitable (but socially needed) train services were put into place, together with a precise definition of the concession contracts and their main terms. With regard to results, in general, most of the restructuring experiences detailed below seem to have been positive. Most countries achieved the objectives of stopping the industry's drain on the state's resources and stabilizing market share for both passengers and freight. Likewise, the companies succeeded in raising their levels of productivity.

Nevertheless, one must take into account two important caveats for future regulation. First, the process of privatization each country chooses depends on the basic objectives sought: to maintain an industry with one operator or a

Table 5.3. *Deregulation and Privatization Experiences in Railways*

Country	*Market structure*	*Ownership of railways*	*Ownership of infrastructure*	*Separation between infrastructure and services*	*Regulatory framework*	*Reasons for deregulation*
Argentina						
Before restructuring	Public monopoly	Ferrocarriles Argentinos (FA), public enterprise with little autonomy	State owned	Unified management under FA	Prices are regulated	High public subsidies, reduce FA's deficits Improve traffic levels Improve productivity
After restructuring	Franchise system for 6 freight and 7 passenger concessions (4–5 operators)	Private companies Operating in each franchise	State network open to third parties	Management of companies Trackage rights exist	Free prices with maximum level Minimal frequencies and quality service	
Brazil						
Before restructuring	Freight: RFFSA Passenger: CBTU	Public companies	Public ownership	Unified management	Regulated prices	Antiquated, inefficient railway industry Reduce state contributions Favor development and regional equilibrium
After restructuring	RFFSA and CBTU divided into sub-networks	RFFSA privatized	Public ownership	Management by the companies Trackage rights exist	Prices control	

(table continues on following page)

Table 5.3 continued

Country	Market structure	Ownership of railways	Ownership of infrastructure	Separation between infrastructure and services	Regulatory framework	Reasons for deregulation
Chile						
Before restructuring	EFE (83 percent of network) and private mining companies (FEPASA)	EFE, public company Private mining companies (FEPASA)	EFE and mining companies	Unified management of the existing companies	Regulated prices	Reduction of state subsidies Improve efficiency of system Increase market share
After restructuring	EFE and subsidiaries, and mining companies	EFE and subsidiaries, private companies	Spread among EFE and other companies	Separation of services and infrastructure on public lines Trackage rights exist	Liberalized prices	
Japan						
Before restructuring	Monopoly (JNR)	Public	State owned	Unified management	Regulated prices	Reduction of state subsidies
After restructuring	6 passenger companies (regional monopolies), 1 freight	Only 3 in process of privatization	Owned by the 6 new passenger companies	Unified management (passenger companies) Trackage rights (4 freight companies)	Free prices	Improve productivity

(table continues on following page)

Table 5.3 continued

Country	*Market structure*	*Ownership of railways*	*Ownership of infrastructure*	*Separation between infrastructure and services*	*Regulatory framework*	*Reasons for deregulation*
New Zealand						
Before restructuring	Monopoly in hands of New Zealand Rail Ltd. (NZRL)	Public agency	State owned	Unified management	Prices and service level regulated	High public subsidies and reduce NZRL's deficits
After restructuring	Monopoly	Private (private groups that bid highest to buy the company)	Lease	Unified management	Free prices	
Sweden						
Before restructuring	Public monopoly	Statens Järnvägar (SJ), government department	State owned	Unified management	Controlled prices	High public subsidies and reduce SJ's deficits
After restructuring	Monopoly on infrastructure and quasi- monopoly in services	SJ, public company with wide autonomy, and presence of small private companies	Managed by a public agency, Banverket (BV)	Separation Services run by SJ and small companies Infrastructure by BV	Control over tariffs has been reduced Not on access prices	Improve traffic levels Improve productivity

(table continues on following page)

Table 5.3 continued

Country	*Market structure*	*Ownership of railways*	*Ownership of infrastructure*	*Separation between infrastructure and services*	*Regulatory framework*	*Reasons for deregulation*
United Kingdom						
Before restructuring	Public monopoly	British Rail, public body with managerial autonomy	State owned	Unified management	Freedom of prices, except in some services	High level of public subsidy Improve traffic and productivity levels
After restructuring	Competition for the market System of 25 franchises in passengers and 2 companies for freight	Private concessions and rolling stock leased to private firms	Private company (Railtrack regulated)	Total separation	Free prices RPI-X in access pricing	
United States						
Before restructuring	Competitive situation	Private companies	Owned by railways	Trackage rights exist	Price control and no closures of loss-making lines	Loss-making companies Loss of markets
After restructuring	Competitive situation with concentration of big companies and many small ones	Private companies	Owned by railways	Trackage rights exist (Amtrak), but 25 percent of trackage with several freight operators	Price freedom and closures of loss-making lines	

Source: Authors.

small number of operators, or to facilitate a process of competition on the track. Second, legacies from the traditional mechanisms of regulation should be avoided. In particular, one must deal with the two common problems of high debt levels and overstaffing before starting any privatization policy.

Experiences in Railway Privatization

The separation of infrastructure from operational services in railways is relevant in this sector, and it conditions the concessioning process in many countries. Vertical unbundling, aimed at solving the natural monopoly issue described earlier, not only promotes greater allocative efficiency, it also encourages some other relevant regulatory questions, which become particularly important as the degree of private participation in rail infrastructure management increases. In this case, the economic regulation of infrastructure should be governed by the adequate combination of three standard principles: fair access to the infrastructure, cost recovery, and efficient access pricing.

ACCESS TO RAIL INFRASTRUCTURE. Regulation of rail infrastructure includes not only simple pricing principles, but also access rights and long-term development provisions. Each country addresses these differently: most have opted to retain infrastructure in the public sector, creating state management agencies (Sweden's Banverket) to regulate private train operators (as in Argentina). Others (France and Germany) have established independent state-owned enterprises to manage rail tracks. Only the United Kingdom has privatized infrastructure and operations.

Whether in public or private hands, rail industry infrastructure regulation must include minimum investment requirements to prevent short-term myopia and ensure that key investments are given priority over increasing dividends or defending against a potential takeover.

Regulation must also address the issue of access, which is particularly relevant in the case of highly integrated transnational networks (as in Europe) or privately or publicly managed dense networks (as in Canada, the United States, and some Asian countries). In the European Union, for example, Directive 91/440 directs each member state to grant international access and transit rights to international groups in which stakes are held by railway undertakings in that or other member states. No directives or resolutions have been related to domestic traffic, although the European Commission advocates the extension of these provisions to all freight and international passenger services. Most countries simply charge (monopolistic) train operators for the use of (public) rail infrastructure.

In the wholly privatized structure of the United Kingdom, open access to passenger services has been limited by a number of provisions that moderate competition. Initially designed to protect rail franchisees from new entrants and from each other, these provisions were anticipated to be gradually reduced over time. In other countries, the contract also clearly specifies access rights, as mentioned later for Argentina, Burkina Faso, and Côte d'Ivoire. In certain large cities, such as Mexico City and Buenos Aires, operators share a common network under a unique transport authority.

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

COORDINATION AND INTERMODAL COMPETITION. A relevant issue when considering the pricing of rail infrastructure is intermodal competition. As mentioned earlier, modal choices can be heavily distorted because of different cost coverage ratios and the use of different cost input bases.[4] A solution is to follow an integrated, multimodal approach. Basic principles will have to apply to all transport operators, irrespective of the mode in which they operate. For example, countries such as Argentina and Chile considered the extent of road freight transport when designing rail concession contracts. The general rule was that operators undertaking business at their own commercial and financial risk should not be at an undue disadvantage in relation to those who enjoy public aid or indirectly benefit from huge externalities.

In the case of natural monopoly infrastructures, the principles envisaged to avoid these distortive effects should be solidified in the coordination of existing networks (particularly in dense rail areas) and the establishment of mechanisms that facilitate interoperability and international

4. One of the reasons for the decline of the rail industry is the fact that road transport did not internalize its social costs (in terms of pollution or safety, for example). More important is the fact that some countries, such as the United States, provide cross-subsidies that benefit heavy trucks.

links. Not even the most advanced infrastructure regulations (such as the Swedish and the British systems), however, offer much help, because they were conceived for a single-country environment. In other countries, such as Argentina before the restructuring process, railways attempted to solve national transport problems by offering underpriced passenger services or subsidized low-quality freight transport. As a result, their financial performance rapidly deteriorated in an isolated framework. Therefore, the infrastructure pricing strategy in these areas should be compatible with the achievement of both local and international objectives, by establishing, if needed, a system of slot assignments in more congested corridors.

VERTICAL SEPARATION. According to Kopicki and Thompson (1995), one of the most clearly defined patterns emerging from deregulation and restructuring is that they carry out two critical dimensions, summarized in table 5.4: the degree of vertical separation between infrastructure and services, and the involvement of private management in the sector.

With respect to the first dimension, the vertical organization of the railway industry has three main options: (a) vertical integration, (b) competitive access, and (c) vertical separation.

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

Alternatively, in the complete vertical separation scenario, the management (and, possibly, the ownership) of facilities is fully separated from other rail functions. This is attractive because, although infrastructure may remain a natural monopoly, it is separated from rail services, where potential competition among different operators is possible. In general, the main advantage of this vertical unbundling is that rail transport is placed in a similar situation as road transport, especially regarding the tariff system and infrastructure planning. Governments could study investment proposals on the basis of a cost-benefit analysis, while pricing

Table 5.4. *Alternative Organizational Structures in Railways*

Private participation	*Total vertical integration*	*Vertical unbundling* *Competitive access*	*Vertical unbundling* *Vertical separation*
Government department	India, China, former socialist countries		
Public enterprise	European railways		
Reformed public enterprise	Many European railways at present		Sweden
Service contract with private sector		Japan (HSR) U.S. (rolling stock) Pakistan (ticket sales)	U.K. (rolling stock)
Management contract with private sector	Nigeria (1980)	U.S. small railways	
Leasing to private sector		Amtrak (U.S.) (track) VIA (Canada) (track) Japan (track) Cameroon (baggage)	
Leasing from private sector		U.S. and Europe (wagons and cars)	
Concession (franchising)		Argentina, Brazil, Chile, Côte d'Ivoire	U.K. (passengers)
Joint venture		Canada U.S. (pipe and wire)	U.K.
Private company	New Zealand	Japan (in progress), U.S. (Class I), Canada	U.K. (freight, infrastructure)

Source: Galenson and Thompson (1993).

policies could be based on social cost.[5] In addition, separating infrastructure from services greatly facilitates the entry of more than one operator on a single route. For profitable services, this would permit notable improvements in efficiency by allowing direct competition among operators. For nonprofitable services, infrastructure separation can be accompanied by tendering to stimulate increased efficiency through competition for the market, to promote the introduction of innovations, and to encourage marketing improvements.

The vertical unbundling of the rail industry, however, also implies several disadvantages. The main problem is the potential loss of economies of scope derived from the joint operation of tracks and services. Often noted is that the relationship between the services supplied and the rolling stock used, as well as the quality, quantity, and technical characteristics of the infrastructure, is so close that both aspects need to be planned together. Thus, assigning different services to several operators may decrease the utilization of the sector's staff and physical assets. Another negative factor is that the new system has a higher risk of becoming less attractive to the user than an integrated system.[6] Also mentioned is that vertical separation requires such a complex institutional arrangement that the resulting transaction costs often are prohibitive for many countries. A final disadvantage of vertical separation is the reduction of investment incentives. For example, an infrastructure owner considering an investment on a facility with only one potential buyer will anticipate bargaining away some of the benefit from the new service once it comes on-line. This problem becomes less relevant with more competition in the market, because competition weakens the bargaining position of individual operators by reducing the specificity of the assets.

PRIVATE PARTICIPATION. With respect to the dimension of private participation in the industry, Galenson and Thompson (1993) provide a list, ordered

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

6. For example, because of the lack of interchangeable ticketing, an integrated national network, and so forth.

in terms of increasing private participation, of the different situations found in the world's rail industry. The first situation is a government department, in which the government fully controls and finances the railroad, which therefore is subordinated to its interests.

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

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

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

New Regulatory Scenarios. The vertical separation/private participation bidimensional space creates a new regulatory framework in the rail sector. It introduces significant new roles and functions for the regulator and modifies the number of possible regulatory structures and models. In practice, choosing a particular method for railway restructuring depends on a number of objectives or goals that the government must balance according to the economic environment in which it operates.

One of the first elements to consider is the existence of financial constraints. If they are important, maximizing the proceeds obtained from the restructuring process will be a primary goal. A second element to consider is the pursuit of internal (or cost) efficiency in terms of providing services at the lowest possible cost, and therefore generating an efficient use of resources. Similar is the goal of attaining external (or allocative) efficiency by setting optimal prices equal to the marginal social cost, which from an intermodal viewpoint, facilitates the best distribution of traffic. The objective of dynamic efficiency requires the long-run minimization of cost through active, technology-improving investment policies. Equity objectives also are possible, such as facilitating transport for all citizens independent of income level. Finally, the government can also consider the optimal allocation of capacity, which favors management of railway capacity, coordination with other modes of transport, and overall minimization of risks in terms of service maintenance over time, risk of default, and so forth. Table 5.5 presents the combination of these objectives, creating at least eight different possible regulatory scenarios, grouped in increasing order of private participation. Not included are some additional scenarios, such as the mixed forms described previously.[7]

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

---

7. This is because many of these forms of private participation are related to very specific services (for example, the case of service or management contracts), and some of the forms of contracting (such as leasing) are occasionally similar to those established in a concession or franchising system.

Table 5.5. Different Rail Regulatory Scenarios and their Objectives

Scenario	*Objectives*						
	Fiscal	*Internal efficiency*	*External efficiency*	*Dynamic efficiency*	*Risk minimizing*	*Capacity allocation*	*Equity*
1 Vertical integration and government department	×	×	✔	✔	✔	✔	✔
2 Vertical integration and reformed public enterprise	×	×	✔	✔	✔	✔	✔
3 Vertical separation and reformed public enterprise	×	×	✔	✔	✔	×	✔
4 Competitive Access and concession regime	✔	✔	✔	unclear	unclear	×	✔
5 Vertical separation and concession regime	✔	✔	✔	unclear	unclear	×	✔
6 Vertical integration and private enterprise	✔	✔	×	unclear	×	✔	×
7 Competitive access and private enterprise	✔	✔	×	unclear	×	×	×
8 Vertical Separation and private company	✔	✔	×	unclear	×	×	×

Source: Authors.

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

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

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

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

Concession Contracts. Despite the number of potential regulatory scenarios just described, few railways around the world have been fully privatized. Instead, most countries have opted to concession rail services, and even rail infrastructures in some cases, to private firms in exchange for a fixed payment. This has been the favored form of restructuring because it allows the government to retain ultimate control over the assets, while the private sector carries out day-to-day operations according to prespecified rules devised in a contract, which transforms the problems associated with traditional regulation into issues of contract enforcement (Thompson and Budin 1997).[8] Because many variables need to be considered, one cannot reduce rail concession contracts into a single standard model. Based on experience, however, table 5.6 proposes six key variables to consider.

The first critical aspect of a concession is determining its type, both in vertical (functional) and horizontal (geographical) size. Recent concessions in the rail industry have created smaller horizontal packages throughout the country. For example, rail freight systems in Argentina, Brazil, Colombia, and Mexico were split into several regional companies, and Chilean railways were broken down into four passenger companies and two freight companies with a separate infrastructure firm. All these countries also used economic criteria to design the size of the concessioning package, accounting for the profitability of different lines. Preferred in Europe is functional separation between infrastructure and services, especially since the promulgation of European Commission Directive 91/440. The privatization of British Rail used this form of concessioning at its most extreme, and also included the private provision and management of rail infrastructures. Sweden and other European countries have developed a less extensive vertical separation, when infrastructure has not been auctioned off to private firms (Lundberg 1996).

The second key issue in designing rail service and infrastructure concession contracts is defining the award process and duration of the concession. This includes the auction rules and, particularly, the criteria defining how each concession will be awarded to a private operator. The award criteria can be chosen from a number of possibilities, for example, maximum payment to government or minimum tariff. One can also choose between unrestricted bidding and bidding that could involve some preselection (Guislain and Kerf 1995; Kerf and others 1997). In the

8. The list of countries with actual or planned rail concessions includes Argentina, Bolivia, Brazil, Cameroon, Chile, Colombia, Congo, Côte d'Ivoire-Burkina Faso (international link), Guatemala, Jordan, Malawi, Mexico, Mozambique, Peru, and the United Kingdom.

Table 5.6. Key Variables in Designing Rail Concession Contracts

Features	*Variables*
Type of contract	Package size depends on economies of scale/scope and existing potential for competition
	Horizontal concessions (geographic) according to country's characteristics
	Vertical concessions (functional) according to network's characteristics (including current state of infrastructure and new investment needed)
	Mixed packages depending on profitability and bidders' financial constraints
	Freight versus passenger concessions depending on relative traffic shares
Award and duration	Prequalification requirements to reduce risks
	Type of auction (sealed, one-shot) and explicit rules for auctioning
	Selection based on government's objectives (fiscal, equity, or efficiency)
	Short periods (favor competition; diminish investment incentives) versus long periods (favor invest ment; diminish enforceability)
	Termination: re-auction preferable to automatic renewal
Contents	Concessionaire:
	• Obligations: services (with adequate performance) and payments
	• Rights: exclusivity and compensation for public service obligations
	Government:
	• Risk sharing (net cost/gross cost mechanisms)
	• Asset ownership

(table continues on following page)

Table 5.6 continued

Features	*Variables*
Price control	Price control depending on monopoly power and social objectives Ideal criterion: marginal cost rules Practical mechanisms: rate of return regulation and price cap schemes Other schemes: price discrimination and cross-subsidization
Quality regulation	Quality of service Safety and externalities Dynamic quality: investments Instruments for quality control
Infrastructures	Access to rail infrastructures Access pricing Coordination and intermodal competition

Source: Authors.

privatization of the former British Rail, for example, the concession process began with a prequalification stage, followed by a formal invitation to tender for a particular package. After indicative bids were received, four bidders were short-listed. One of these was subsequently named the preferred bidder, and was given two weeks to complete financing and other organizational arrangements before being confirmed the winner. At that point, the regulator gave public details of the bid, in terms of the required subsidy and promised service improvements.

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

Although the guiding principle should be to maximize competition so that the most efficient firm ends up winning the award, clearly no single method stands out for selecting the winner once bids have been submitted. The final choice depends on the government's objectives, which should be explicit and built on transparent criteria. Thus, if the government intends private participation to be a means of reducing the burden on the public sector, it must use fiscal benefits as the main criterion, looking at who requires the lowest subsidy or who offers the highest auction price. For example, Brazil successfully auctioned the six regional rail concessions to the highest bid above the government's minimum price. Concessionaires were required to make an up-front payment immediately after the auction, followed by a stream of predetermined payments over the life of the concession. Similarly, in Britain, minimizing subsidy payments appeared to drive the choice of bidders, especially in the first concessions. Other criteria were the financial position of the tenderer, its managerial competence, and its operational proposals.

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

The rail industry has awarded many concessions using formulas with multiple criteria, which can account for a larger number of objectives. For example, in Argentina, the bids for the six freight packages that were concessioned were evaluated using the net present value of the canon to be paid to the government during the first 15 years of the concession, the quality of business and investment plans, the staffing levels, the proposed track fee for passenger trains, and the share of Argentine interest in the consortium. The weights of these criteria reflected both the importance attributed to investment in the railways and political compromises on employment. For the award of metropolitan commuter railways, however, the Argentinean authorities kept things simpler to make the bidding process and final selection as transparent as possible. They learned from the freight concession that selecting the winning bid through numerous cumbersome criteria with discretional weights was more likely to reduce the efficiency of the bidding process than to improve it. Instead, the terms of the concession should be made clear to all potential bidders, and bidding should take place on the basis of a single parameter encompassed in the bidders' economic assumptions in terms of the concession.[9]

With regard to the optimal duration of the concession contract, the trade-off is evident in terms of efficiency, because the shorter the concession, the more immediate the competitive pressure, but the less the incentive to invest and develop the business. Longer concessions, in contrast, tend to diminish the regulator's enforcement capacity and soften the incentives to promote efficient outcomes. The general rule is to adapt the concession period to the economic life of the assets and to make this compatible with the government's objectives. This balance often creates conflict: while concessionaires generally argue for long contracts that provide them with incentives to build up the business and purchase or replace long-lived assets,

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

concessioning authorities prefer shorter lengths to favor the achievement of efficiency (by the implicit threat of nonrenewal) and fiscal goals (because the canon or auction price may be increased after the first few years of the concession). Only if sunk investments are minimal, and asset reutilization is possible, are shorter periods advisable for particular rail services (those related to signals, track, and station maintenance).

Shaw, Gwilliam, and Thompson (1996) point out that the average duration of a rail service concession is 5 to 10 years, increasing up to 30 when network investment and development are included. In Argentina, for example, the six freight packages were concessioned on a 30-year term, with an optional 10-year extension, due to the poor state of infrastructure and the huge investment that was required. For similar reasons, the international rail link between Burkina Faso and Côte d'Ivoire was awarded in a 15-year concession. Conversely, train operating companies in the United Kingdom were granted a concession to run passenger services for a period of only 7 to 15 years.

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

Because renegotiation costs money, but a lack of renegotiation might cause performance deterioration, concession contracts should specify the circumstances for renegotiation, and which party should initiate the process. If intermediate objectives are achieved, a prescheduled revision process might help to reduce both parties' risks. Although the contract will always be incomplete, standard clauses should include behavior in unforeseen changes in demand conditions, responses to unanticipated rises in energy or labor costs, and so forth. For example, in Argentina, freight concessionaires could not fulfill their promise to invest US$1.2 billion in the rail network over 15 years due to unexpected falling traffic levels.

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

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

These operations may include infrastructure provisions if they were auctioned off to private firms. In fact, a large part of railway activities might be concessioned. These include infrastructure such as track, signals, stations, yards, and shops; operating equipment such as locomotives, wagons, and carriages; and general service access to track, route, and schedule information and maintenance. The exact form in which this process is developed in practice depends on the parties' risk-sharing agreements. According to a service contract, for example, train operators provide rail transport services for passengers or (rarely) freight according to specific routes, levels of quality, and technology as established by the regulator. The operators may cover some investment costs and carry some commercial risk, which can be integrated into a net cost contract, in which the operator keeps all revenues generated by passenger or freight traffic. This type of contract, in which the operator carries revenue as well as cost risk, often generates more traffic and is let to the most attractive bid, but it offers a higher incentive to predate. Alternatively, gross cost contracts specify that all revenue accrues to the government and that the contracts are let on the basis of the least total cost supplier, so operators carry cost but not revenue risk. The experience in the United Kingdom with regard to passenger franchises suggests that gross cost contracts generate more bids per tender (particularly from new entrants), offer greater incentives to public revenue generation, reduce the administrative cost for the regulatory authority, and support any fare scheme with modal integration and quality control.

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

The implementation of rail concession exclusivity rights varies in each country. In Argentina, freight concessionaires have exclusive use of tracks but must grant access to passenger operations in return for a compensatory track fee. In Chile, passenger services and infrastructure initially remained in public hands, while freight services were privatized. The 15-year concession for the Côte d'Ivoire–Burkina Faso transnational railway was awarded with a 7-year exclusivity period, after which the operator should grant access to third parties specified by the regulator for an agreed-upon fee. Thus, exclusivity rights should be viewed as another instrument for regulatory control, and not taken for granted by the firms ex ante. Limiting the duration of the monopoly period balances the regulator's desire to reap the benefits of competitive access to the tracks with the private train operator's preference for full control of the market to generate profit and facilitate revenue forecasting. In general, most railways have been concessioned on an exclusive basis in geographical areas, as in Argentina or Brazil, possibly with some access rights for connecting railways to certain central or strategic track segments. This has been due to the geopolitical configuration of the country, the density of the existing network, and the need to promote competition in major markets (as in Mexico) or for noncompeting services (such as passenger services on freight tracks in Chile).

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

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

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

Price Regulation

According to standard economic principles, prices for rail transport services should match the opportunity cost of providing them so as to make the most efficient use of the economy's resources. This is the economic efficiency or first-best criterion, which has defined the traditional regulation of the rail industry during the past 50 years. The main focus of government regulation was to control market power by setting prices that limited the monopolistic abuse of any particular railroad. The exact form of tariff control (official approval of rates with little or no degree of financial autonomy) in each case depended on the nature of the industry, the ownership of the assets, the complexity of the regulated service, and the social and political pressures to maintain financial equilibrium in the medium and long run.

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

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

In the new regulatory environment in which separating infrastructure from services can be relatively easily achieved, and a notable degree of private participation in rail management exists through concession contracts, pricing principles must be put into practice by means of concrete rules within the contract. Because rail concessionaires are now able to set prices relatively freely, the concession contract should include a procedure to control the prices set by operators. One should generally set these price control mechanisms according to three key factors: (a) the degree of monopoly power effectively conferred to the operator; (b) the extent of government noncommercial objectives in the concession award procedure; and (c) the possible existence of limiting factors, such as intermodal competition. This latter element is relevant in rail freight operations (intermodal competition from trucking),[10] but in the case of passenger traffic (especially commuter and regional), social pressure for low fares usually dominates many price interventions. In practice, the most common alternatives (second-best criteria) for price control in rail concessions are rate of return regulation and price cap mechanisms.

Rate of Return Mechanisms

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

---

10. For example, in Argentina, railways only carried 8 percent of total freight ton-kilometers at the time of concessioning.

Revenue requirement = total cost = (variable cost) + (rate level x base rate)

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

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

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

Despite its advantages within the traditional price regulation mechanisms (mainly its simplicity), three additional problems are associated with this sort of regulation. First, it gives little incentive for productive efficiency, because firms can pass production costs on to final users in the form of higher prices; second, it leads to excessive investment and capital use because the firm is guaranteed a return on investment; and, finally, the high degree of discretion the regulator enjoys in determining the base rate and the rate of return reduces the incentive for rent-seeking behavior by the regulated firm. This is the so-called Averch-Johnson or capital bias effect, which is not particularly adverse in developing economies whose capital needs are seldom fulfilled.

Price Cap Regulation Mechanisms

The most common alternative to the standard rate of return regulation is using cost-plus incentives that, in practice, take the form of a menu of cost reimbursement rules that firms themselves select according to their preferences for sharing operating costs with the regulator.[11] These mechanisms basically aim to achieve dynamic efficiency (in the sense of the regulated firm achieving the lowest unit cost in the long run) by sharing some of the efficiency improvement rents between the firm and the regulator.

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

Price cap regulation is another incentive that both railways and other privatized utilities use. In its most standard form, it consists of setting traditional maximum price schemes based on long-run marginal costs in order to offer a firm an incentive to achieve the goal of dynamic efficiency while maintaining all or part of the gains associated with the firm's future increases in efficiency. This mechanism came as a consequence of the criticism directed at the lack of cost minimization embedded in rate of return regulation and other traditional price regulation mechanisms. One has to balance its efficiency gains, however, with the higher information rents that it implies (De Rus 1998).

14. See Guasch and Spiller (1996) for detailed examples in other industries.

The price cap system has a number of minor variations. In the rail industry, one of the most developed is the RPI-X formula. In this setup, the price for a basket of the firm's prices can increase in any one year by no more than the increase in the retail price index (RPI) for that year, minus some fixed-cost (efficiency-related) parameter *X*.

$$price_{\ (year\ 1)} \leq price_{\ (year\ 0)} \times (RPI - X)$$

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

The United Kingdom, for example, has applied the price cap mechanism, in its RPI-X formula, to passenger traffic franchises. Commuter fares are regulated with respect to a basket containing all relevant fares, weighed broadly by the income that the operator derives from each. For three years from January 1996, increases in the capped fares were not permitted to be more than the retail price index increase from the 1995 base price; after January 1999, the price cap was planned at RPI-1 percent.

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

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

One of the most critical issues is the setting and resetting of the productivity X-factor. A possible method consists of using indexes or indicators (as described below) to measure the difference between aggregate rates of growth of outputs and inputs, and therefore calculate productivity from the residual. Econometrics also provides alternatives for estimating cost functions and their corresponding productivity parameters (see Borts 1960 for a classic reference). Once the X-factor is determined, the initial price ceiling imposed on the firm after a switch of regime is critical. If the caps are too high, then too little surplus is transferred to consumers and deadweight losses are huge. If they are set too low, the firm may not be able to break even and may then have difficulty attracting capital, leading to a deterioration of service quality.

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

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

These methods are valid not only for limiting monopoly profits earned in passenger or freight traffic, but also in controlling infrastructure access prices (discussed later). Finally, because tariff controls can easily be cheated on quality grounds, quality requirements become essential for monitoring overall performance of rail concessionaires.

The Problem of Rail Infrastructure Costs

As described earlier, rail infrastructure provision and management are characterized by a high ratio of fixed to marginal costs, the existence of avoidable costs, and unavoidable or common costs. Avoidable costs are uniquely associated with a particular output: if this output is not produced, no cost is incurred. This guiding principle relates to the idea of cost recovery for particular outputs. Avoidable costs may thus be considered as a floor to regulated prices (if any), because charging less than the avoidable cost is equivalent to operating at an economic loss. This makes standard pricing rules inoperable in this sector, because first-best or efficient principles of marginal cost pricing may result in large deficits that jeopardize the long-run survival of the firm. Three particular problems then arise with respect to allocating the rail infrastructure costs: cross-subsidization issues, cost-recovery problems, and the possibility of setting inefficient prices (Talley 1988).

Illustrating the cross-subsidization problems in pricing rail services or infrastructure produced in the presence of common costs is the case of a profit-regulated railroad connecting two large cities and also providing rail service to a smaller town along the route between the two cities. The fares charged for passage from the small town generate revenues exceeding the additional cost of serving it, such as ticketing and station costs, but not sufficient to cover an equal or proportionate (however defined) share of the common costs, such as trackage, signaling, and train yard costs. The issue is how to allocate common costs among customers and services. In many cases, cost sub-additivity and efficiency require joint production and allocation of fixed costs among all services, without cross-subsidization (accounting for externalities whenever present).

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

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

The cost recovery principle should be a central issue in the design of any rail infrastructure pricing procedure. The theoretical and political debate focuses on two options. Many public firms still advocate the use of the efficient price mechanisms described earlier in this section and propose marginal cost rules with the simultaneous use of public subsidies to cover fixed costs. Alternatively, a growing literature patronizes the use of full-cost recovery prices, including price discrimination, multiple-part tariffs, or cross-subsidization schemes, if needed. Although one considers the possibility of it yielding inefficient outcomes for the theoretical efficiency principles, it constitutes the second-best available alternative in most cases.

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

Many econometric studies have shown that in the case of the rail industry, the marginal cost of those railways that are still vertically integrated lies in the range of 60 to 70 percent of average cost; where rail services are separated from infrastructure, the marginal social cost of rail infrastructure alone often is well below the 60 to 70 percent range (Friedlander and others 1993). Price discrimination, if feasible and politically acceptable, may help to raise cost recovery to around 60 percent of total cost without driving demand off the market. Thus, full cost recovery would require a further price markup of more than 60 percent above the efficient price. Economists have defended alternative proposals, in terms of the so-called Ramsey pricing principle, for infrastructures with high fixed costs and low marginal costs.[12] They rarely work in practice, however, because they arouse consumers' suspicions of unfair treatment and undue discrimination. Moreover, under Ramsey pricing rules, all unattributable fixed and common costs are apportioned on the basis of the services' demand characteristics.

---

12. Ramsey pricing refers to charging higher prices above unit costs to more inelastic market segments. When infrastructure and services are separated, their use becomes more complicated and still is not clearly solved, because one must estimate different demands for services as well as for tracks.

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

The Access Pricing Problem

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

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

Depending on the discretion allowed to the integrated firm, one can determine potentially distortive effects on access prices in several ways. First, when infrastructure is still publicly owned or managed, the regulator can determine the price as an integral part of the access terms defined in a contract with one of several private train operators. Second, the regulator may allow the firm to choose from a menu of alternative regulatory schemes, usually rooted in incentive-based price regulation mechanisms (to favor the firm that achieves higher levels of efficiency). Third, the firm may have discretion over aspects of access pricing subject to some overall regulatory constraint. Finally, the firm may have full discretion over the price and only be restricted by the country's antitrust law.

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

The first problem to be solved is that of the actual value of capital assets: nominal value versus potential to generate cash. While the latter is clearly a function of the privatization and regulation methods and the extent of competition envisaged in bidding for the right to operate concessioned infrastructure services, the former is more likely to reflect a past situation that domestic reforms are trying to overcome.

The second method of setting access prices consists of developing usage-related charges. Once-avoidable costs are covered by increasing prices that are inversely related to demand elasticity. Another, less controversial, option is using a two-part tariff to avoid service cuts by train operators to save charges even when the network has no cost saving.

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

Railtrack owns the rail network and sets track charges upon which the rail regulator must agree under the criteria openly published in a number of regulatory policy statements. The price control system operates through a simple RPI-X formula that is revised every five years, remaining fixed between revisions. For example, in January 1995, the regulator announced the price controls that would apply to franchised passenger services from April 1995 to April 2001.

The structure of Railtrack's access charges for franchised passenger services is based on the usage-related charges made up of multiple-part tariffs that have at least four elements.[13] First, track usage charges tend to reflect short-run effects on maintenance and the renewal costs of running trains of different types for different distances. Second, traction current charges recover the costs of electric current, varying geographically and temporally and reflecting distance covered and type of vehicle. Third, the long-run incremental cost indicates the long-run costs imposed on Railtrack in delivering the total access rights of a train operator. Finally, the remainder of the fixed charge are common costs, designed to recover the rest of Railtrack's costs at the subzonal, zonal, or national level. This is apportioned among train operators on the basis of budgeted passenger vehicle miles for subzonal costs and budgeted passenger revenue for zonal and national costs. The first two elements amount on average to only about 9 percent of total track access charges, and given the current structure of charges, these are the only elements that vary directly. The remaining 91 percent of the aggregate charge is in the form of a fixed charge, which does not vary with the number or type of trains run or with passenger revenue.

In the case of freight services, access prices are more flexible. The rail regulator has simply established several principles for Railtrack to consider in its relationship with private operators. First, prices must cover the avoidable costs Railtrack incurs as a direct result of carrying that particular freight flow; second, prices must be lower than the stand-alone cost that a national efficient competitor would incur; third, no undue discriminatory charges are possible; and finally, charge structure should reflect the value to users of access to the rail network and enable Railtrack to recover its total cost.

---

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

As opposed to the British case, the setting of access charges in other European countries is still underdeveloped. In 1995, the European Union passed two directives concerning the application of Directive 91/440 on the separation of infrastructure management and transport operations. Directive 95/18 regulated the licensing of railway undertakings, and Directive 95/19 established several general principles on allocating railway infrastructure capacity and infrastructure fee charges. These principles were designed to ensure an optimum, nondiscriminatory use of infrastructure and guarantee an access charging policy according to European Community rules, but member states received them with varying degrees of enthusiasm. The objective of most governments that have set rules for infrastructure fees is to cover costs and differentiate fees to reflect different cost factors.

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

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

Quality and Safety Regulation

Quality performance is important when society evaluates the economic contribution of the rail transport sector to the social welfare. The particular level of quality that train operators achieve, and the particular features of three main dimensions that broadly define quality in the rail industry (service, externalities, and investment), critically determine the value added by this transport mode. The first questions that naturally arise are why quality regulation is needed at all in this industry, and to what extent this regulation relates to the standard price regulation mechanisms described in the previous section. Economic theory provides a well-known argument to answer these questions: real-world transport activities are characterized by market failures due to information problems.

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

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

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

Definition of Quality Targets

In setting up the quality standards incorporated in concession contract designs, the regulator often uses the principles of yardstick competition.[14] One may construct these quality standards at the national or regional level with inter-industry comparisons (as in Brazil and Chile for many of their public utilities) or by establishing international benchmarks or best practices (as in Australia for transport services and infrastructures).

One considers three elements in detail when designing this process. First, as in other transport modes, quality is mainly measured in concrete service levels or specified service standards. This measurement, however, is suited more for factors such as train punctuality, the reliability of aboard services, and the waiting time at stations or platforms than it is for other factors.[15] Simultaneously, the services provided before the transport itself, such as ticketing, reservations, and luggage or cargo handling, are often ignored as part of the rail industry's value chain, although they may constitute relevant aspects of both intramodal and intermodal competition. For these reasons, the first element to consider in designing a quality control in the rail industry is an integrated vision of transport service that includes not only the ride itself, but all aspects related to infrastructure (track and stations), stations, and pre- and post-transport services provided to clients.

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

---

14. This is done to avoid the problem of regulator's capture and the discretionary nature of the regulatory action. Making undue comparisons between different rail systems, however, is a risk.

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

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

Intermodal coordination with other industries is also necessary, because social quality performance is always evaluated in relation to feasible alternatives. Saturated corridors (where investment in roads, railways, and airports clearly overcomes demand) are a waste of resources that few economies can assume. This almost general equilibrium approach to evaluating quality constitutes the third element of the quality regulation process, although, in this case, it is not particular to this industry. The sociopolitical implications of quality regulation (in terms of equity or public service obligations and the social acceptance of quality standards) determine the overall quality targets to be established in each industry.

Taking into account these three characteristics, table 5.7 summarizes the five most important quality dimensions for the railway industry (vehicle, route, service, social, and dynamic quality) along with a number of standard performance measurement instruments for them. The first three (vehicle, route, and service) are related to what is usually named quality of service, whereas the last two refer to static and dynamic externalities.

QUALITY OF SERVICE. Different countries have dealt in different depths with regulating the quality of rail transport services in regard to vehicle quality, the transport service itself (aboard trains), and the pre-and post-transport services, although, as described earlier, a positive correlation exists between the extent of the restructuring activity in the rail industry (in terms of private participation or separation of infrastructure from services) and the quality regulation requirement imposed on the industry post-restructuring.

In general, countries in which the sector is still heavily dependent on government or public agencies (such as in Asia and Eastern Europe) have done less to establish separate quality control frameworks than those in which private participation has been significant (such as the United Kingdom) and detailed quality control systems have been set up. In all cases,

Table 5.7. Quality Dimensions of the Rail Industry

Dimension		Definition	Measurement variables
Quality	Vehicle of service	Aboard quality (wagons, locomotives)	Age of vehicle/number of years in service Vehicle size and load factor Availability of seats Accessibility Travel comfort • noise • vibration • temperature • tidiness
	Route	Route quality (travel of passengers and cargo)	Distribution and numberof stations Timetable • peak trains • first-last train • weekend-commuter services Frequency (number of trains perhour) Punctuality/reliability (waiting at stations) Cargo services (reliability)
	Service	Pretransport and post-transport service quality(added value to service)	Ticket sales/reservations Handling Staff adequacy and competence Inquiries and general information Response to complaints
External quality		Externalities (safety and environment)	Public service obligations Safety procedures Liability regimes Environment protection (noise, pollution) Congestion
Dynamic quality		Investment policy	Fleet and track renewal rates Track and stations maintenance Investment obligations

Source: Authors.

the basic principle governing the design of quality mechanisms is that customer service should be paramount if railways are to maximize profitability and compete with alternative modes of transport. The economic relationship between separate units in a railway enterprise should be structured to ensure the preservation of incentives for maximizing customer service (Swift 1997a,b).

This is particularly relevant to the separation of infrastructure and operations. Vertical unbundling in railways distances infrastructure management from the end-user customer and could yield undesirable side effects or contradictions. For example, the density of traffic (trains per day) that maximizes returns on infrastructure investment is likely to be greater than the optimal level from the operators' point of view. This is because at high densities, passenger service is likely to suffer due to congestion. Therefore, no matter whether the separation is institutional or only financial, one must incorporate mechanisms to compensate infrastructure units that run below optimal capacity into contracts to maximize end-user customer performance as a whole. Because the particular characteristics of the rail industry in each country require fine-tuning of any regulatory or contract enforcement mechanism, table 5.8 proposes a simple scheme that identifies and separates the roles to be assigned to the regulator and the operator (either franchisees or public or private monopolies) with regard to quality of service regulation.

After the reform of the United Kingdom's rail system and the full privatization of its services and track provision, that system constitutes one of the most practical examples of a detailed quality of service regulatory framework (see table 5.8). For example, in the case of passenger transport, the regulatory agency (Office for Passenger Rail Franchising, or OPRAF) defines what level of service is tendered for particular routes and corridors and sets the minimum level of service for every route in the country (not only timetable specifications, but also journey time, first and last

Table 5.8. *Role Assignment in Railway Quality of Service Regulation*

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

Source: Authors.

departure times, and so forth). If franchises operate a poorer service than specified, then OPRAF reserves the right to withhold the grant.

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

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

In practice, one cannot incorporate all the quality dimensions defined in table 5.7 in the same proportion to any service quality mechanism. The British system mainly focuses on the route dimension and is based on its extensive experience with deregulation. When the role assignment that table 5.8 proposes is not considered, or its components cannot be easily separated, several quality regulation failures may arise. The most important is the failure to define adequate independent quality measures. This is the case of several rail concessionaires in Argentina, where the level of vertical integration between the train service providers and the maintenance firms (in the form of subsidiaries or units integrating a larger industrial group) has distorted the incentive to provide the optimal price-quality ratio in favor of more frequent repairs and technical updates.

SAFETY AND EXTERNALITIES. Regulating the quality of service is only one of the two static aspects of quality regulation to be considered in designing a global framework for quality regulation in the rail industry. One also must consider the social or external dimension of quality regulation, including all issues related to safety and externalities (such as pollution and congestion), and it specifically differs from level of service quality regulation in at least four aspects.

The first element is the scope of regulation. Because noncompliance with social quality standards may affect users and nonusers of transport services, these standards should always be exogeneously set, by national or supranational legislation with intermodal implications, in the case of the rail industry. This is not always the case for timetables, load factors, or vehicle size, variables that usually have simple intrafirm consequences. In the European railway industry, for example, one can find three levels of quality regulation. Directive 91/440 determined the overall principles, and the obligation to comply was envisaged in mode-specific regulation (for example, the Railways Act in the United Kingdom) or in legislation that applies to all sectors of the economy (for example, the Health and Safety Act).

The second factor that makes service quality regulation different from social quality regulation in the rail industry is that one must use a regulatory approach in the latter. Because the risks associated with accidents or potential environmental damages not only directly affect the private benefit, but also the social benefit of this transport mode, an external regulator or agency is needed to coordinate safety and reliability. This coordination is particularly important when firms move from a public to a deregulated system, as described earlier. Furthermore, in the rail industry, separation of infrastructure from services and the introduction of open access have made it necessary for a rail track controller to ensure safe coordination between different operators that are using the same tracks or stations.

Again with the British railway system as an example, the safety regulator is the health safety executive, who informs and advises the Office of the Rail Regulator. Operators of railway services, stations, and networks must have an accepted safety case before the office approves their license. A safety case is a complete resource, control, and management plan for delivering safety and defining safety procedures, organizations, and systems. The private infrastructure provider, Railtrack, must have its own safety case, a fundamental component of which is Railtrack's Safety Management System, which is a system of operational and technical standards to ensure safety and safe interworking in Railtrack's infrastructure.

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

Despite tragedies in 1998 and 1999, railways traditionally have a good reputation for safety, a perception that converges with statistical proof in most countries. Therefore, one could conclude that safety levels and management are quite sufficient and no particular safety precautions or measures should be taken. Public outcry, negative social effects, and adverse public opinion from a single catastrophe, however, together with the persistence of regular fatalities (staff accidents, passengers joining and alighting trains, and so forth), make it impossible for the regulator to avoid designing measures and policies for diminishing individual and social risk.

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

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

For example, air pollution is one of the most regulated areas in the road and air transport modes, but is not a critical issue in the rail industry, however, with a few notable exceptions in certain countries and routes. Noise pollution in suburban neighborhoods, areas close to stations and depots, and delicate countryside ecosystems has attracted more attention from both the public and regulators. Most countries, therefore, incorporate into their regulation the design and specification of measures to reduce noise produced by rolling stock and stationary sources (fans, compressors, and generators) and shunting noise.

The final issues related to environmental regulation are measuring, analyzing, and predicting the emissions of chemical substances (heavy metals, lubricants, dust, and so on) where railway lines are present and assessing the risk of rail-related activities (such as transport of dangerous goods) to the safety of local residents. In these cases, most countries subordinate their social quality standards and the role of their regulators to the overall technical principles emanating from supranational organisms or professional associations. Private and public rail transport operators are obliged to comply with national and supranational environmental standards. Europe, for example, has European Community directives on vehicle air pollution that specify environmental standards for vehicle engines and fuel qualities that apply to both vehicles (wagons, locomotives) and transport operations.

DYNAMIC QUALITY: INVESTMENTS. Table 5.7 lists a third dimension in quality regulation of the rail industry. Because the regulatory process is by itself a dynamic relationship between the regulator and the regulated transport providers, firms, and passengers, one must take into account the dynamic links in this relationship when certain quality standards are controlled. In particular, the investment policy of the railroads is the most important dynamic element to be considered in the design of concession contracts, particularly with respect to the implications that these investments will have on the future performance of the firms. A complete quality regulation regarding the investment policy must first define who decides the investment objectives in terms of fleet and track renewal rates, track and station maintenance, and future investment obligations. When the regulator assumes this role, it must also set up adequate mechanisms to monitor the progress of the investment stages, and provide the incentives for avoiding abandoning projects before they reach conclusion.

When the regulated rail transport provider is in charge of its investment policy (with respect to the renewal or maintenance of its fleet, for example), a quality control should also be imposed to avoid, for example, inadequate planning or excessive unnecessary repairs as a means of earning extra revenues from subsidiary companies. Some countries, notably the United Kingdom, exert this control by isolating noncommercial investments and investment planning and making them the subject of specific public grants.

One of the most controversial issues in the concession contract is the relationship between this investment and the prices set to recover it, because ex ante prices are decisive in determining the extent and mix of investment in new rail infrastructure. Uneconomic investment decisions have historically been imposed on railways, which in most countries has been the main cause of accumulated debt. As described earlier, insulating railway operators from such debts has been a central aim of the railway reform and restructuring processes. Thus, in principle, the simplest decision according to standard economic principles is to proceed with projects whose net present value, calculated according to a suitable discount rate, is positive. In theory, the most obvious discount rate to use in public sector projects is the interest rate on long-term government bonds.

In practice, however, this bond rate may not be appropriate in several circumstances, and governments that choose a discount rate lower than this rate invariably find they cannot proceed with all of the projects with positive net present value.[16] This means that in certain cases, specifying a hurdle rate (which determines whether the project will in fact be implemented) to test against the project's internal rate of return is more useful. Therefore, on pure economic efficiency grounds, if the selective process is strict and calculations are correct, only projects that do not generate losses on new investments should proceed, in order to avoid later problems with cost recovery relating to infrastructure pricing. On social grounds, however, few rail investments would pass this strict cost-benefit analysis, and subsidies to pay for the fixed costs may be required at the investment point.

For investment financing, the more important consideration from the viewpoint of quality regulation in the concession contract design is the monitoring of the operator's financial health to prevent possible cheating incentives (for example, lowering the quality of building materials in tracks, signaling mechanisms, or stations). In principle, rail investments do not have specific criteria according to which the regulator should impose particular rules with respect to the firm's capital structure. When the size of the investment is large enough, the private concessionaire seeks the adequate mix between debt and equity that enables him or her to carry out the project. Only if there is government or other public participation in the new investment, should the concession contract regulate the conditions and terms under which the asset transfer (if any) takes place.

Alternatively, when the main (or sole) source of funding for the infrastructure provider is revenue from track charges, one uses different criteria to reflect two basic circumstances: to maintain existing standards or to increase capacity and quality of service. In the former, the regulated access

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

charge is set to provide the regulator's cost of renewal. The performance regime provides the incentive for the infrastructure provider to undertake investment. When expansion or improvement of capacity is required, the track manager is expected to finance investment with increased revenue, so that train operators and infrastructure providers share both the risks and benefits of an improved infrastructure.

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

Instruments for Quality Control

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

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

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

The third mechanism that can be used to control quality is a customer compensation scheme, in which grants or payments are awarded to people affected by noncompliance with quality standards. In practice, these mechanisms only work if one can easily verify quality failures. This requires a detailed regulation not only of quality standards, but also of monitoring rules and guarantees for both the regulator and the regulated that the inspection process will be transparent and objective. Moreover, if the compensation is distributed to consumers, either directly by the firm or through an intermediary body, sharing rules must be also defined. The practical difficulties associated with this quality control mechanism have led many countries to instead specify minimum quality standards for certain parameters of the rail industry, backed by explicit legal sanctions that may include fines or the revocation or withdrawal of the operating license.

Finding the adequate mix of these control mechanisms is often the most difficult task in designing the quality regulation process. Table 5.9 outlines most countries' approach, with a summary of the most important instruments.

In conclusion, the quality regulation process consists of three stages. First, before entry into the market (stage I), the goal is to anticipate and minimize future conflicts between the regulator and the concessionaire.[17] Licenses must specify the expected characteristics of the service in terms of, for example, routes and frequencies of trains or timetables. For passenger services, particularly in the case of urban and suburban trains, one can also set vehicle capacities and punctuality. Finally, so as not forget the dynamic dimension of quality described above, stage I must also specify investment plans and financing rules. Afterward, during market operation (stage II), instruments for quality control in the rail industry should mostly be related to the direct monitoring of the firm's performance. Thus, this is the time to introduce quality incentives in price mechanisms, to establish the firm's obligation to reveal information and the auditing (external or internal) processes to be carried out. In most cases, using technical control instruments (such as tacographs or track electronic controls) complements the standard instruments. Finally, after the transport activity has already occurred (stage III), the regulator can implement compensations or punishments according to any of the schemes described above. Both penalties and incentives must be graded according to the expected future evolution of the relationship, because severe fines or large subsidies may alter the behavior of the operator in the market.

17. To achieve this, one can use pretender qualification requirements to ensure a minimum level of technical and practical expertise and financial solvency, as described in the previous section.

Table 5.9. Instruments for Quality Control in the Rail Industry

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

Source: Authors.

Performance Indicators

The rail industry uses performance indicators to monitor the behavior of one or more regulated firms to evaluate the effectiveness of the regulatory measures to which they are subjected.[18] The main advantage of these indicators or indexes is that they provide a periodical assessment and control of the firm's activity and continuously update information, simply, quickly, and at a relatively low administrative cost for the regulator.

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

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

Similarly, simple indicators should be carefully interpreted over time to avoid contradictions and inappropriate measurements. For example, in the assessment of railway output, the number of trains/kilometer may be relatively high, while passengers/kilometer or tons/kilometer may be relatively low (if the firm specializes in one type of traffic). Given this conflict, overall performance can be ambiguous. The most practical solution is to jointly interpret the indicators and the objectives that they serve. For example, a service quality objective, such as the number of trains per hour, may conflict with both financial objectives, reflected in a high cost recovery rate, and objectives based on the maintenance of low prices.

18. For example, one can establish quality indicators, as defined in this section, in a contract and review them regularly to confirm that the terms of the license are being fulfilled.

Thompson and Fraser (1996) point out that monetary and productivity variables should be carefully defined for interfirm comparisons. Fares, wages, outputs, and inputs vary widely among countries for many reasons that are not necessarily related to the firm's operations, but to measurement or statistical errors. For example, average passenger fares are based on the overall mix of passenger classes (each with a different price). Tariffs are often higher per passenger/kilometer for short trips than for long ones, and they must also depend on the existence of government subsidies or artificial compensations. Similarly, common freight tariff mistakes include not accounting for the different mix of commodities, size of shipment, or length of haul. The latter also affects passenger traffic and is particularly relevant because some costs (ticketing, billing, and station maintenance, for example) are fixed with respect to the length of the trip but vary with size or distance.

These difficulties are increased when measuring productivity, because a simple comparison among partial measurements of output cannot capture the complexity of relationships or the variety of productive structures that take place within a rail operator. For example, a commonly used productivity indicator, the number of passengers-kilometer or tons-kilometer per employee, depends on such diverse factors (for example, regulatory environment, structure of the labor market, availability and quality of infrastructure, or alternative transport modes) that it could be seriously misleading if interpreted without care. The term employee can refer to terminal staff, administrative staff, train crew, or maintenance staff. Similarly, capital can be disaggregated into trains, wagons, terminals, platforms, routes, and so forth.

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

Main Types of Indicators

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

Several external factors that vary widely from country to country and firm to firm substantially influence comparisons. Contextual indicators assist in comparative analysis and define desirable performance levels. They include social and economic characteristics of the railways as well as other elements associated with the economy as a whole. Directed mainly at the regulator, they control for the exogenous factors in interfirm and intrafirm comparisons. Table 5.10 presents several examples from international statistical sources.[19] Simultaneously, many indicators (particularly those for prices and quality of service) are informative to transport users and provide input for the regulator's control tasks. Jointly with the contextual indicators, these management indicators provide the necessary instruments for judging the management and behavior of the company, and one can group them at three different levels, summarized in table 5.11.

Some final practical rules that could be helpful in this process are as follows: (a) each indicator should have at least a function or objective; (b) the relationship between each indicator and its objective must be clear and direct, although (c) multiple indicators (jointly interpreted) can address multiple objectives; and finally, to assure the utility of the indicators, (d) appropriate data must be provided and (e) the management of the indicators' information should be part of the regulatory process.

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

Indicators of service quality that were highlighted earlier should serve the same way as price indexes to establish evaluations of different companies, as well as dynamic or time evaluations. These measurements should be analyzed together with price indexes because of the possibility of finding different feasible combinations of price and service quality. For example, a high number of trains per hour—in other words, a high traffic density—could only be financed by means of high prices.

---

19. In particular, the International Union of Railways publishes an annual summary of the main statistics of its affiliated railways, although not all of them are always available for all railroads.

Table 5.10. Contextual Indicators in the Rail Industry

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

Source: Authors.

The simultaneous implementation of control systems for prices and service quality may limit the firm management and reduce operability. Placing an emphasis on price control or service quality depends on whether the regulator prefers services at the lowest possible price or services with

Table 5.11. *Management Indicators in the Rail Industry*

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

Source: Authors.

certain standards of quality. All these indicators allow the regulator to monitor the operators' activities as defined in stage II of table 5.9. Unjustified or systematic breaches of quality standards (insufficient number of trains per hour, lack of punctuality, unreliability, very high indexes of load factor, and so forth) should be accompanied by an appropriate system of penalties, as described earlier.

Best Practices

Taking into account the above comparison caveats, the remainder of this section compares some of the most relevant (or most desirable) performance indicators for the rail industry with the best results actually achieved. Table 5.12 shows this procedure, which many governments around the world use as a yardstick mechanism. The last two columns show the best practice values for a sample of European, Australian, and American rail companies and the values considered desirable according to a World Bank study by Gannon and Shalizi (1995) and the Australian Bureau of Industry Economics.[20]

One of the most useful insights that these examples can provide is the clarification that setting desirable values for indicators is a difficult task. One should put extreme care into making exclusionary comparisons. For example, according to figures, the unit revenue ratio has a desirable value below US$0.04 in passenger traffic and US$0.03 in freight. Lower values, such as those in several European countries (0.036 and 0.019, respectively), could indicate lower prices or low fare collecting efficiency. In either case, regulation in each country will notably affect the prices charged by each company.

The measurement for average train speed should distinguish between passenger and freight transport and among their different categories (urban, regional, long-distance, international, and so forth). The desirable indicator estimates an average speed of 60 to 100 kilometers/hour, but in each country it should depend on the type of traffic, the social and economic level of the country, and the relative importance of the railway in its development.

Similarly, different regulatory policies, as well as other variables such as vehicle size and journey type, influence the measurement of the average passenger load in terms of the number of passengers per train. In Europe, the Italian national company attained the highest level for 1994, with an average of 197.5 passengers per train over the year. Correspondingly, a European railway also attained the highest level of passenger

20. Every year, the Australian Bureau of Industry Economics publishes a benchmarking report (see BIE 1995, for example) that compares its main utilities (including rail transport) worldwide.

Table 5.12. Best Practices in Railway Management Indicators

Indicator (example)	*Best practice*	*Desirable*
Commercial		
Revenues and prices		
Passenger revenue/passenger-km (in US$)	0.036	0.04
Freight revenue/ton-km (in US$)	0.019	0.03
Freight to passenger tariff ratio (percent)	—	—
Commercial services: general		
Average train speed(in km/h)	—	60–90
Arrivals with small delays (10-15 min.) (percent)	96	90–95
Ratio of lost plus damaged freight (percent)	1	1
Commercial services: passengers		
Number of passenger per train	197.5	—
Passengers-km per route-km (total)	5237	—
(in thousands per km)	(136)	—
Commercial services: freight		
Number of tons per train	604.13	—
Tons-km per route-km (total)	2,819.19	>2,000
(in thousands per km)	(352)	—
Operational		
Labor productivity		
Passengers-km per employee	—	—
Tons-km per employee (in thousands)	11,000	>750
Passenger trains-km per employee	—	—
Freight trains-km per employee	—	—
Total trains-km per employee	4,434.84	—
Capital productivity		
Availability of locomotives (percent)	914.28	—
Tons-km per wagon	—	>80%
Wagons-km per wagon	—	—
Freight and passenger wagons availability (percent)	—	>90%
Financial		
Cost coverage		
Costs covered with total revenue (percent)	—	>100
Costs covered with typical revenue (percent)	—	>80
Cost reduction required to reach break-even (percent)	—	<0

Note: Desirable values are only approximate and should be taken as general references that might vary across countries and regions.

Sources: BIE (1995); Gannon and Shalizi (1995).

traffic density—5,237 passengers/kilometer per route/kilometer, for the Dutch operator. For freight traffic, the equivalent figures were 604 tons per train, for the Finnish national operator, and 2,819 tons/kilometer per route/kilometer, for the corresponding Belgian company.

As mentioned above, the measurement of productivity is often grouped around the labor and capital indexes. Because many companies do not detail by activity the volume of employees, however, usually only the aggregate index of total trains/kilometer per employee is available. Previous studies (Nash 1985) have estimated that freight traffic is more labor-intensive than passenger traffic, so this measurement is clearly biased due to the different composition of the output of railway companies. Considering the aggregate index indicative of the volume of trains/kilometer per employee in Europe, the most efficient company in 1994 was the Dutch operator, with a volume of 4,434 trains/kilometer per worker. In North America, where many companies offer only freight transport, the most efficient companies transported about 11 million tons/kilometer per employee.

One can divide measurements relating to the productivity of capital into those that refer to traction units, locomotives, or wagons. For locomotives, an interesting index is locomotive availability (percent), which indicates the degree of overdue and deferred maintenance, for which Gannon and Shalizi (1995) recommend a value not less than 80 percent.

Finally, financial indicators should not be less important to a regulator, even one who is more concerned with operational and commercial performance. For example, the ratio of revenues to total costs may indicate the degree of financial solvency, whereas the level of subsidization and subsidies as a percentage of total revenue or costs indicates the degree of financial dependence on public bodies. These indicators are very important and should not be independently interpreted, because, as shown earlier and by empirical evidence (see Gathon and Pestieau 1995), the most heavily subsidized railway companies are often the most inefficient.

The main conclusion is that performance indicators are useful but should be designed and interpreted with care. Reference levels and comparisons are only provisional guides and are not normative. Individual indicators must not be analyzed in isolation from others. A unique optimum does not exist for any indicator, nor is there an optimal profile for several. Their appraisal requires trade-offs that measure the relative cost of changing different indicators and the relative importance of the objectives that the indicators reflect.

Conclusions

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

6

Toll Roads

Antonio Estache, Manuel Romero, and John Strong

Road transport has long been, and will be for a long time, the dominant form of transport for freight and passenger movement throughout the world.[1] In Latin America, for instance, road transport accounts for more than 80 percent of domestic passenger movements and more than 60 percent of freight movements—more than 85 percent in some countries such as Argentina and Brazil. In Africa, the proportions are even higher. Not only is the sector large, but it is still growing rapidly in many parts of the world. In Asia, from 1984 to 1994, the road networks of Indonesia, Korea, Malaysia, and Pakistan grew in length by more than 5 percent per year. In Eastern Europe, countries historically dominated by rail are now witnessing a rapid expansion in the demand for road transport. In Russia, the total freight moved by road is expected to increase from just over 10 percent to almost 40 percent within the first decade of the collapse of the former Soviet Union.

Toll Road Services

Because most road projects require investments with slow amortization periods, and many of these projects will not generate sufficient demand to

Notes: For additional information and more data, see the World Bank web site on toll roads at www.worldbank.org/html/fpd/tranport/roads/toll_rds.htm. Part of the discussion in this first section draws on J. A. Gomez-Ibanez' "Pricing," in Gomez-Ibanez, Tye, and Winston 1999.

1. See Heggie and Vickers 1998, chapter 2, for more details on the overall role of the sector.

make them self-financed through some type of user fee or toll, the road sector continues, and will continue, to be in the hands of the public sector to a much larger extent than the other transport activities.[2] However, fiscal crises and competing demands from other sectors such as health and education are bringing changes in the extent of public-private partnership in the expansion and operation of road networks. Governments throughout the world, including many poor African and South Asian countries, are commercializing their operations to cut costs, improve user orientation, and increase sector-specific revenue.[3]

The search for increased private sector participation in the road sector applies to national, high-traffic roads (at least 15,000 vehicles per day is a good bet for viable tolling of roads) as well as to roads falling under the responsibility of all government levels. Many urban roads, often under the responsibilities of subnational governments, are now also facing strong increases in demand. From Argentina to Thailand, Australia to Canada, major arterial roads are being built under toll road concession schemes. Because in many of these countries the governments are increasingly finding it difficult to finance the costs up-front, they are giving the private sector concessions to construct and operate these urban roads for a specified period of time before inheriting these assets at the end of the contract, typically at a zero cost. Toll roads (publicly or privately managed) can represent a large proportion of the high-traffic highway systems (up to 80–100 percent in some countries), but they generally represent a very small share (5–10 percent) of the total paved road network (but up to over 30 percent in such countries as Argentina, France, and Korea).

To be effective, toll road projects must meet many requirements to ensure that a regulator's implementation and monitoring of these concessions is smooth. This chapter focuses on the lessons of the international experience with toll road privatization and regulation. It is oriented somewhat more toward project finance than the other chapters, to illustrate the importance of contract design for a regulator. An effective contract design at project time

2. In addition, pricing decisions in this sector tend to be influenced in many countries by strong trucking lobbies that aim to keep cost recovery as slow as possible.

3. These partnerships not only aim to convince the private sector to finance its investment needs but also to participate in reforms to cut costs in its operation and maintenance. This chapter, however, focuses on toll roads. For a more detailed discussion of the commercialization of the sector and of other forms of public-private partnerships in the sector, see Heggie and Vickers (1998) and Ecole Nationale des Ponts et Chaussees (1998).

is crucial, because often there is little more for a road concession regulator to do than monitor that all involved parties comply with their contractual commitments. Furthermore, in many ways, the renegotiation of a contract is often a replay of the initial negotiation with a different distribution of information between the regulator and the concessionaire. Having full grasp of the basics of contract design in project finance is required for regulators.

While toll roads in specific settings seem to be in demand, the problems that many of this first generation of road concessions have met, from Mexico to Thailand, have given toll projects a poor reputation. Many mistakes were made. What is obvious is that tolling is not the best solution for every road. One can design a project in many ways to get the private sector involved without having to toll the road (see box 6.1). Most of these alternatives aim at improving efficiency (in other words, lowering costs). But many ways also exist for getting the private sector involved in toll roads and thus reducing public sector financing requirements. Understanding the context in which toll roads are viable is necessary both for their initial success and for their effective long-run regulation.

The Broad, Relevant Economic Characteristics of the Activity

When considering an increased role for private activity in the road sector, the most immediate challenge to confront is the enormous range in the development, quality, and performance of the sector in any given country. These varied settings create different operating and investment requirements and hence potentially very different types of possible packaging to make roads attractive to private investors. This section summarizes the most relevant stylized economic facts surrounding the core decisions to make in designing toll road packages.

THE OVERALL SUPPLY OF GOOD ROAD SERVICES IS STILL LIMITED. The overall market potential is good because the unsatisfied demand for good road services is great. Moreover, the need for improvements in networks is, in many countries, at least just as great. Indeed, many of the new public-private partnerships are for road upgrading and paving rather than for greenfield projects. Even if 100 percent paving of existing roads is unlikely to be a realistic target for many countries, the margin for improvements suggests a reasonably good market for private road operators and interested construction companies. Indeed, in developing countries, the proportion of the main road network that is paved averages to a modest 45.5 percent (it varies from a low of around 2.5 percent to a high of 100 percent).

Box 6.1. *Contracting Out Road Planning and Management*

Road agencies are increasingly contracting out:

- The planning and management of selected roads to consultants and contractors
- Entire road networks
- Donor-financed small infrastructure projects.

Argentina, Australia, New Zealand, and the United Kingdom are using the first model. The United Kingdom started the process in 1986 when its Department of Transport (DOT) decided to package parts of the motorway network into commissions and then invited bids from consultants to take on the responsibility for maintaining all roads and related structures within the commission to a prescribed standard. The winning consultant then organizes a competitive term contract between the owner (DOT) and the contractor, which then carries out all work on instruction from the consultant. In one of the largest commissions (West Yorkshire, with 330 lane-kilometers, 305 bridges, 420 kilometers of drains, 950 road signs, and 3,400 lighting columns), costs fell by well over 15 percent, the consultant took on 29 of the 34 DOT staff who were made redundant (one moved to another job and four took early retirement), and quality and flexibility of the maintenance regime increased.

The second model involves contracting out the management function for the whole network under the jurisdiction of a selected road agency. Industrialized countries generally do this to increase efficiency and as part of the redefinition of the government's role. Developing countries mainly do it to ensure that a competent body that remains answerable to the local district council manages small urban and district roads. Some small municipalities in the United States at the county council, and at the district level in the United Kingdom (where it is called externalization), use this model. Zambia also uses it for both urban and rural district councils. These arrangements offer great potential for dealing with small road networks.

Francophone Africa uses the third model extensively. The AGETIP is a contract executing agency (like a private sector project implementation unit) set up to execute donor-financed infrastructure projects. The agency generally has a board composed of well-known figures (which do not include government representatives), a general manager appointed by the board, other line managers (an administrative and financial manager and a technical manager), and staff hired under private sector terms and conditions of service who are paid competitive salaries. The agency is set up as a private, nonprofit association and pays no taxes. It works on behalf of local authorities who delegate certain functions to the agency. The local government usually reserves the right to select the projects, and the agency then (a) recruits consultants to carry out detailed engineering; and (b) invites bids and awards contracts for supervision and works, manages the contracts, and pays the contractors directly from a special account opened in its own name. The agency is subject to bimonthly management and financial audits and an annual technical audit.

Source: Heggie and Vickers (1998).

Furthermore, private operators can also easily improve road maintenance. In Latin America, for instance, over the past 15 years or so, most governments have spent roughly one-quarter of what they should have spent to maintain roads. This is why new private operators finding out that rehabilitation costs are higher than expected is not uncommon when they take over the responsibility from the government. This is often a subject of dispute between regulators and private concessionaires once private operation has been in place for a while. Overall, from a strategic viewpoint, contracting out road improvements can provide a smooth phase-in or learning process into the development of a future relationship with potential private investors on new roads.

THE DEMAND FOR HIGH-TRAFFIC ROADS IS STILL GROWING. The demand for road services will continue to grow, and hence so will the need for investment. Worldwide, the stock of motor vehicles is growing at nearly 3 percent per year. Because the number of vehicle-kilometers traveled tends to grow somewhat faster than the stock of motor vehicles, this implies that, at least for some segments of the road network, the demand prospects are quite good. The FAST urbanization of the developing world adds another dimension that cannot be ignored and explains the strong demand for urban access roads in many of the most populated countries of the world. The challenge here is betting on the right horses. Demand will increase, but only on some segments of the network, and a government may be tempted to oversell a specific road based on aggregate traffic growth prospects.

Even with the effects of toll levels held constant, traffic volumes are sensitive to income and economic growth. The failure to recognize this may be one of the main reasons that so many toll road projects have failed or ended in bitter renegotiations. Motorization and vehicle-kilometers traveled tend to increase faster than income levels. This high income elasticity, especially for leisure trips, makes toll roads especially sensitive to macroeconomic conditions. For roads that serve export activities, exchange rate changes can dramatically affect trade, leading to major changes in demand patterns.

Many toll road projects in the past decade have dramatically overestimated traffic levels. In some of the Mexican road concessions, traffic volumes were only one-fifth of the forecasted levels. In Hungary, the M1 Motorway attracted only 50 percent of its expected volume in its first year of operation. The Dulles Greenway, outside Washington, D.C., only attracted a third of its expected daily volume. Even after a toll reduction of 40 percent, the Greenway still was only able to achieve two-thirds of its originally forecast volume.[4]

4. The Dulles Greenway experience suggests a toll price elasticity of –2.3, a very high sensitivity. This result is due in part to the upgrading of a parallel alternative route. Other estimates range from –1.4 to –2.5, quite income elastic in all cases.

THE DEMAND FOR SAFETY IS GROWING AS WELL. Investment needs and types also will have to address the need for improved safety. Each year, more than 700,000 people are killed and more than 10 million are injured in road accidents, costing the global economy about US$500 billion. About 70 percent of these accidents take place in developing and transition countries, where road accident rates per 10,000 vehicles tend to be 20 to 30 times higher than in industrialized countries and cost up to 2 percent of gross domestic product. This is a major concern for policymakers and is becoming a concern for private roads operators as well, because policymakers tend to include requirements in concession contracts that address the need to drastically increase highway safety. A peculiar problem arises when governments change safety regulation in the course of a contract, thereby implicitly changing the investment requirements and hence the terms of the agreement with the private operators.

THE DEMAND FOR A NETWORK OFTEN DOMINATES THE DEMAND FOR A SPECIFIC ROAD. The demand for a toll, and hence the risk attached to a road, often depend on the fact that the toll roads have to be built into integrated networks. Reformers often forget that the tolled part of a road network benefits tremendously from the existence of a public road network around it. In practice, the value of a specific road depends a great deal on the extent to which it benefits from a complement of public and private roads. More specifically, the network characteristics of the sector mean that benefits from investment at one point in the system can depend on service flows and capacities at other points. This implies that both public and private roads operators need to take into account a number of service obligations.

THE MARKET FOR TOLL ROADS IS SENSITIVE TO GLOBAL FINANCIAL CONDITIONS. The financial crises that affected many emerging markets in 1996–99 had a dramatic effect on the evaluation of toll road projects. Project sponsors and creditors experienced difficulties due to macroeconomic factors, and financing became much more costly and of shorter term, thereby adding refinancing risks. Many toll road projects that were required to generate returns of 15 percent in the early 1990s have been reevaluated in light of project experiences to date and because of macroeconomic uncertainty. The result is that required returns for toll road projects appear to have risen to 20 percent or more. The key factors are cost, average daily traffic volumes, and the willingness to pay tolls. Overall, average daily traffic volumes in excess of 10,000 vehicles per day seem to be required to attract private capital. Below this traffic volume, various types of government support, such as grants or guarantees, are likely to be required. The effect is to make many proposed concessions nonviable, or at least to cause their deferral until greater corridor demand is assured and a more stable financing environment exists.

A project's ability to obtain financing, however, is not solely determined by its underlying cost and demand. The country and concession environment and the nature of public-private risk management also have important effects on the viability of toll road programs. A stable economic and political context has been essential for a sustained toll road program. Because toll roads typically are high-performance highways, they are particularly dependent on income levels and economic activity. Moreover, because toll roads also tend to be politically visible, they may be subject to attempts to influence project selection, implementation, and operation, especially through attempts to delay tariff increases and to evade toll collection entirely.

The Specific Economic Characteristics of the Activity

National characteristics are important in developing a greater private role for the road sector. At the same time, the privatization teams must clearly understand the economic characteristics of each toll road package to design an appropriate public-private partnership. These economic characteristics are determined by a number of factors, including the project's function, its physical characteristics, and the underlying market demand.

THE PURPOSE OF A TOLL ROAD. The project's economic characteristics should be the starting point for designing the appropriate role for the private sector. The first question to ask is: why is the toll road being put in place? In many instances, tolling is being considered for fiscal reasons. Governments often want new, stable sources of finance, and regulators must be aware that this can influence tremendously the choice of a toll road design and pricing form, as discussed later. One can classify toll roads as congestion relievers, intercity arterials, development roads, or bridges and tunnels. Their main characteristics are typically categorized as follows:

- *Congestion relievers* are relatively short roads built to relieve traffic on existing urban routes. SR-91, for example, expands capacity of a major highway in southern California in the United States.[5] Congestion

5. One can express capacity in terms of the variable passenger car unit (PCU). A PCU is a measure equivalent to the space occupied by a car, so a coach that occupies approximately twice the space of a car corresponds to 2 PCUs and a truck to 2.5 PCUs.

relievers, while expensive to build because of land costs, generally have significant revenue potential because they tend to serve heavy traffic demand. The high land acquisition and construction costs, however, may require high tolls if privately financed, so pricing decisions and regulatory oversight become important. In addition, because congestion may be concentrated at peak periods, time-of-day and other variable pricing schemes may be required. Tolling is becoming more widely used as a mechanism to manage traffic demand on increasingly congested highways, a change made easier by advances in tolling technology that have made tolling more efficient and more convenient.

- *Intercity arterial roads* are built to improve access between major cities, to airports, or to port/terminal complexes. An example is the Malaysian North-South Expressway, linking the Thai border through Kuala Lumpur with Singapore. These roads tend to be expensive because they are generally long, high capacity, and built to serve heavy truck traffic. Tolling decisions between different types of user groups are particularly important for these roads.
- *Development roads* link more remote areas with urban centers or with major transport routes. An example is the Chilean South Access project that links a forestry region to the port of Concepcion and the Pan-American Highway. While development roads can provide a stimulus to economic growth, traffic volumes generally are not financially sufficient in the early years, and thus these are seen as speculative investments that require substantial public participation.
- *Bridges and tunnels* are typically short, expensive to build per kilometer relative to roads, and, in most cases, serve high volumes of traffic. They are often built as congestion relievers and may have a similar strong financial capability due to traffic volumes. Examples include the Rio-Niteroi Bridge in Brazil and the Dartford Bridge outside London.

THE COSTS OF A TOLL ROAD. Once reformers clearly recognize the purpose of the toll road, its costs must be identified. A project's physical characteristics are the primary determinants of its costs. Important aspects include whether the project is a new facility or an expansion of an existing road; the length, capacity, and design; geographic and geologic aspects; and toll collection mechanisms. New facilities are more costly per kilometer than expansions or rehabilitations of existing facilities. Rehabilitation and expansion typically require less construction work than new facilities.

Moreover, expansion projects that involve preexisting tolled facilities may be able to use the toll revenues to lower external financing requirements. For example, the Buga-Tulua expansion project in Colombia, which connects three major cities, was able to use existing tolls for about a third of project costs. Wider roads (number of lanes), their thickness and construction technique, and the type of geography traversed also are key determinants of project costs. As a result, project costs can vary over a wide range. The South Access project in Chile, which featured favorable geography and mostly rehabilitation work, cost about US$0.2 million per kilometer. By comparison, the Guangzhou-Shenzen highway in China, which involved six lanes through a region subject to flooding, cost more than US$15 million per kilometer. Bridges and tunnels, because of design requirements, tend to be much more expensive; the Dartford Bridge cost US$247 million for 2.8 kilometers, or US$88 million per kilometer (Mercer 1996). The proposed Colonia Bridge connecting Argentina and Uruguay was forecast to cost in excess of US$22 million per kilometer (over US$800 million in total costs).

Finally, recognizing that road capacity presents high levels of indivisibility is important. For example, each lane in a highway typically represents a maximum offer of 2,000 vehicles per hour, but it also represents the minimum offer per lane. If demand is about 3,000 vehicles per hour, capacity will end up being 4,000 vehicles per hour and the market will have excess capacity. So while a bus or train company can adapt the number of vehicles to fluctuations in demand, road service offers full capacity at all times. Therefore, if capacity is designed for peak periods, the road will be underused during off-peak periods. This means that if investment in road infrastructure capacity is carried out for long periods, schemes must be introduced across the board in the toll design to recover investment cost while preventing motorists from being overcharged.

THE DEMAND FOR A SPECIFIC TOLL ROAD. Demand considerations determine the next crucial component of the economic picture of a toll road. One can measure market demand in terms of actual or expected traffic levels, predictability of expected traffic, and the willingness to pay tolls. All these measures are critical to the design of toll road projects, because they determine whether the revenue stream is large enough and predictable enough to obtain financing. The markets served, the number and quality of competitive alternative routes, and the toll road's links to the rest of the transport network also affect traffic levels. Predicting traffic levels is especially difficult for two reasons. First, new projects are unable to rely on existing traffic volumes as the basis for demand forecasts. As a result, they must turn to other methods of demand estimation, such as stated preference models, which may be less reliable.[6] One must make judgments about the new road's ability to draw traffic from existing alternatives and to generate new traffic. The second reason is that in cases in which projects are to be stand-alone, the level of tolls required to cover costs and provide required financial returns may be far above existing toll levels, if tolls are levied at all. In these cases, estimates of price sensitivity and willingness to pay for new facilities become very hard.

THE WILLINGNESS VERSUS THE ABILITY TO PAY FOR A ROAD. Often forgotten is that road investments must take into account the need to serve different user groups, including very poor users in rural areas who may not be able to afford the toll levels required to allow the operators to recover their investments. This is important, because road investments are irrecoverable or sunk in the sense that, once built, they cannot be converted to other uses or moved elsewhere.[7] While investors must be guaranteed the fair opportunity to recover their investments, when preparing privatization, the government must consider the ability to pay of all segments of the concerned population to avoid future tensions between users and operators. This is why the political challenge of introducing tolls is different for greenfield and rehabilitation/upgrading projects. For a given contract duration, tolls for new roads will often be much higher than for rehabilitation projects, because amortization costs tend to be much higher for greenfield projects.

More generally, the experience of Latin America and Eastern Europe shows that the standard assumptions that toll road users are willing to pay high tolls to compensate for reductions in travel time and vehicle operating costs are not as realistic as many would like them to be. This is a major problem, because the tolls that users in these regions are willing to pay may not be high enough to attract private equity (or debt, for that matter). Some practitioners argue that standard traffic models used to forecast the demand for the roads are too mechanical and do not recognize well enough

6. For a review of demand estimation methodologies, see Small and Winston (1999, pp. 11–56) or Trujillo, Quimet, and Estache (2000).

7. The fact that road investment is sunk, rather than subject to economies of scale, is an important distinction. Highway operating costs (with the possible exception of costs imposed by heavy axle loadings) tend not to be very sensitive to volume.

the behavioral changes that toll brings about. For instance, Piron (1999) reveals that for a series of toll road projects in France, the traffic forecast models had omitted a number of critical factors. These included the relative importance of using the toll for the overall budget of the facilities' private or commercial users and the change in the user's willingness to pay with the distance covered.

Privatization and Regulatory Trends

The trend toward increased tolling of roads is clear. The precursors were in the United States and Europe. In the first half of the 19th century, private toll roads outnumbered public roads in the United States. But during the late 19th and early 20th centuries, the growth of the railways and problems with toll evasion led to a decline in private toll roads. Toll road development in the United States further slowed after 1956, when the Federal Highway Act established a federal gasoline tax to fund the interstate highway system and prohibited tolling on new, publicly financed highways. By the late 1980s, though, public funding constraints and infrastructure demands stimulated new interest in toll roads, mostly as congestion relievers in metropolitan areas.

European countries have had more experience with toll roads in recent decades, but with mixed results. Toll financing developed in Europe after World War II because of rapid growth and budget constraints. France used public toll financing in the 1950s and early 1960s, while it introduced private toll concessions in the late 1960s and early 1970s. Only one in four of the French concessionaires have survived, however. Spain introduced private toll financing for intercity motorways in the 1960s; 9 of the 12 original concessions continue to have a major role in Spain's road network. In Italy, more than 20 concessionaires have built more than 5,000 kilometers of toll roads. The largest of the Italian concessionaires, Autostrade, operates most of the highway network. Austria, Denmark, Greece, Norway, and Portugal also widely use toll systems. The Norwegian system is unusual in that it uses concession companies to collect tolls, while the government road administration retains responsibility for design, construction, and maintenance.

The Movement toward Privately Financed Toll Roads

The latest wave of toll roads is in developing countries, where economic and population growth and growing links with international markets led to pressures for more highways. Mexico launched perhaps the most ambitious program of new roads, to build more than 5,000 kilometers of new roads between 1989 and 1994, the majority of which have not met projections and have had to be restructured with significant public contributions. Expansion of existing toll road systems has met better, although still mixed, success in other Latin American countries, most notably Argentina, Brazil, and Chile. China, Colombia, Ecuador, Hong Kong (China), Hungary, India, Indonesia, Malaysia, Peru, the Philippines, and Thailand also have pursued private or public-private toll concessions. Many of these projects are discussed in the context of particular issues in the sections that follow.

In terms of numbers, one of the publications monitoring the development of infrastructure projects, identified 121 projects in developed countries between January 1985 and October 1998. The average project value was around US$750 million (driven by a number of EEC-sponsored megaprojects in Europe). A World Bank database for developing countries identified 280 roads projects in partnership with the private sector between 1990 and 1997. The average project size was around US$190 million, but with a large dispersion across regions. Eastern European projects have reached enormous proportions while South Asian projects have tended to be the smallest. As table 6.1 shows, the bulk of these projects were in East Asia and Latin America.

Table 6.2 summarizes the scope of toll road provision in selected countries. While toll roads are typically only a small share of the total road network, they tend to be located in the most densely traveled corridors and thus have the potential to play major roles in the transport network. Toll roads in many countries comprise a dominant share of the expressway

Table 6.1. *Divestitures, Concessions, and O&M Contracts in Developing and Transition Economies, 1990–97*

	Africa	*East Asia*	*Eastern Europe*	*Latin America*	*South Asia*	*Total*
Number of transactions	5	102	2	93	6	208
Value (million US$)	426	18,567	1,086	18,794.8	63.5	38,937.3
Average project size (million US$)	85.2	182	543	202	10.6	187

Source: World Bank Private Participation in Infrastructure database.

Table 6.2. *Tolled and Other Roads in Selected Countries*
(km)

Country	*Total road network*	*Total expressway network*	*Tolled road*	*Tolled roads (percent of total)*	*Tolled roads (percent of expressway)*
Argentina	500,000	10,400	9,800	1.96	94
Brazil	1,980,000	—	856	0.04	—
Chile	79,800	—	3	0.00	—
France	966,000	14,886	6,305	0.65	42
Hungary	158,600	435	57	0.04	13
Indonesia	260,000	530	530	0.20	100
Italy	314,360	6,444	5,550	1.77	86
Japan	1,144,360	15,079	9,219	0.81	61
Korea, Republic of	77,000	1,880	1,880	2.44	100
Malaysia	94,000	1,702	1,127	1.20	66
Mexico	303,262	5,683	5,683	1.87	100
South Africa	525,000	1,440	825	0.16	57
Spain	343,200	7,194	2,255	0.66	31

— Not available.
Sources: Heggie and Vickers (1998); PadeCo (1999).

network and thus may play particularly important roles in urban areas and in intercity trade.

Experiences with "Privatization"

Many toll road projects have been undertaken, each with different design and investment demands and political and organizational arrangements.[8] Many toll roads have been negotiated quite loosely and have often been the outcome of informal agreements between the government and a construction company. Other programs have been overly ambitious and have resulted in partial or total failures because they were implemented too quickly. Sound toll roads require good planning. The government should consider funding preliminary studies that demonstrate public commitment, increase the future regulator's knowledge base, and help reduce the costs of delivering road services. These studies might involve such matters as environmental and land acquisition needs, indicative traffic and revenue projections (which are essential preparation for both contract design and renegotiation), and

8. This section draws on the World Bank toll road web site.

project design criteria. Design specifications can range from virtually no public sector responsibility for road features to detailed specifications with respect to route, alignment, capacity, locations of interchanges, materials, pavement, and so forth. A lower level of public involvement allows the private sector to provide potentially innovative solutions and better match infrastructure provision to market demand. Allowing this flexibility reduces the ability to compare proposals, however, because different bidders may take different approaches to project design. Projects that have limited opportunities for innovation should be more explicit in design and award criteria.

Experience suggests that reformers should address three key project selection and design issues early in the concession process: whether a free parallel road should be required, the feasibility of cross-subsidies, and whether concessions should be for a single road in a network or for a package of roads.

Requiring Free Parallel Roads. While the idea of having competition between roads is a good one in principle, the evidence so far suggests that traffic levels in most developing countries cannot sustain duplication from free alternative routes. Toll road traffic in such cases has generally fallen well below projections. The Mexican toll road program illustrates the challenges imposed by parallel roads. Launched in 1985, this program introduced a toll road development plan with a range of conditions, one of which was the provision of a free alternative parallel route. Traffic predictions for the concessioned roads suggested that trucks would form about 20 to 45 percent of the traffic. They turned out to be only about 5 percent. A black market in toll receipts was developed by truckers who used the parallel free roads yet produced toll receipts for their employers in order to reap financial benefits. This problem was overcome when the road operators agreed to exempt trucking companies from paying a toll, but this damaged the financial viability of the toll road.

This experience suggests that the competition argument is difficult to implement in an environment where traffic is not strong enough and lobbies are powerful. The best argument in favor of free parallel roads is one of social equity, to ensure that the poor can still have access to the road network, but this often detracts from the new toll road's effectiveness in alleviating congestion and may also cause problems for cost recovery if the toll cannot produce enough revenue. In general, tariff differentiation, as discussed later, will be a much better solution to help the poor, therefore reducing the case for a parallel toll-free road.

Using Existing Concession Revenues to Fund New Projects. In some cases, existing roads have been tolled in order to provide revenue for the construction of new segments in the network. The French pioneered this technique in which new roads with higher construction costs are supported by operating surpluses from existing toll roads (Papon 1998). The Japanese also have been committed to this concept, having used tolls to generate revenue for road construction since the mid-1960s.[9] In 1972 they introduced a toll revenue pooling system. The pools are separate for urban expressways and for regional networks. Tolls are set equally on all routes and segments of the network, no matter what the construction costs or traffic levels. The Japanese felt that traffic forecasts could only be achieved if the full network was in place, and that profitability of some routes would be improved by the opening of connecting routes. Politically, establishing common tolls across the network was easier because it avoided confusion and was fair, because all roads provide essentially the same service. More generally, creating profitability to fund a new road or concession is common. Similar stories can be told about several Asian toll roads programs.

While these examples illustrate that tolling can assist in releasing funds for new construction, from a regulator's viewpoint, the standard risks implied by cross-subsidies require close monitoring of the cost structure of the various roads to ensure that the average toll is not higher than it needs to be. Monitoring the transfer of resources from one group of consumers to another is important, because those who are paying tolls on the existing road are thereby paying for the construction of a new road, which would otherwise have been funded by taxpayers and will provide benefits for other future users. This may be part of a government program of regional development that needs to be explicitly recognized. Toll roads are often developed in congested corridors of a capital city because good revenue streams there are easier to predict. Where this is the case, the investment in the road is benefiting more affluent areas of the country. If this crowds out other investments in less affluent areas of the country, then other regional equity issues are raised.

SHOULD CONCESSIONS BE PACKAGES OR INDIVIDUAL PROJECTS? The project economics of toll roads suggest that traffic volumes must be in the range of 10,000 to 15,000 vehicles per day for toll revenues to be sufficient to cover construction, operating, and financing costs. In many countries, only a few

9. Operating and maintenance costs and interest costs on the construction loans took up 57 percent of the total pooled toll revenue on the 6,416-kilometer National Expressway Network in 1997. Approximately 50 percent of routes generated revenue in excess of their operating, maintenance, and interest costs (Matoba 1999).

such corridors exist. For other roads that may serve important transport roles, bundling a package of roads into a single concession may be possible. The pooling of existing roads reduces the volatility of overall concession cash flows and may thus increase financial viability. In some cases, this may involve transfer of an existing toll road or major untolled route that may need upgrading or expansion, along with an associated feeder network. If properly designed, the feeder network could serve to enhance the viability of the main toll road.

Organizational Options

One can design a road concession in many ways. Table 6.3 presents a spectrum of alternatives for involving the private sector in the provision of toll roads, ranging from maintenance contracts through full build-operate-transfer (BOT) concessions and corridor management. Each option is described in terms of the nature of public and private involvement and typical features (such as duration and project size). The principal responsibilities for toll road development include design, maintenance, toll collection, financing, and legal ownership. In practice, however, governments seldom follow a pure strategy and end up combining various types of contractual arrangements (illustrated by Argentina's restructuring experience, discussed in box 6.2).

Most widely used is the BOT model. This structure can be broadly defined to include variations such as build-own-operate-transfer, build-lease-transfer, rehabilitate-operate-transfer, and similar arrangements that are used to develop new facilities or rehabilitate existing roads. Under the generic BOT model, a private consortium receives a concession to finance, build, control, and operate a facility for a limited time, after which the facility is transferred back to the government. What makes the road sector so special in this context is that in most countries, the consortium includes a major foreign and/or local construction company mostly interested in the short-term use of its assets (essentially machinery) and skills. This often has an influence on the way in which contracts are drafted and also on the speed of the investments to be made. Governments should pay close attention to ensure that investments are driven by demand rather than by the short-term concerns of a consortium motivated by the opportunity for construction profits—as has been the case in too many toll roads projects.

The consortium typically assumes primary responsibility for constructing the project, arranging financing, maintaining the road, and collecting

Table 6.3. Characteristics of Organizational Options for Toll Roads

Features	*A: Maintenance management contract*	*B: Turnkey*	*C: Operate and maintain*	*D: ROT*	*E: BOT*	*F: Corridor management*
Definition	Maintain	Design and build	Maintain and operate	Finance, rehabilitate, maintain, and operate	Finance, design, construct, maintain, and operate	Finance, design, construct, maintain, and operate Develop corridor/ network
Examples	New South Wales Chile, Brazil	United States Hong Kong (China)	Argentina Hong Kong (China)	Argentina, Colombia	• Malaysia, Philippines, Thailand • Argentina, Mexico	U.K. (DBFO) Colombia, Brazil
Direct cost recovery from users	No Payment from government to operator	No Fixed payment from government to operator	Some degree of toll revenue sharing with government	Concessionaire may pay government or vice versa	Government investment usually required Ex-post subsidies not uncommon	Government contributes existing roads and other investment usually required
Scale of private investment	Very low	Considerable for very short term	Low	Medium	High	Medium/high
Private sector risks	Maintenance	Design Construction	Traffic and revenue levels Political Financial	Rehabilitation Traffic and revenue levels Political Financial	Design Construction Traffic and revenue levels Political Financial	Design Construction Traffic and revenue levels Political Financial

(table continues on following page)

Table 6.3 continued

Features	*A: Maintenance management contract*	*B: Turnkey*	*C: Operate and maintain*	*D: ROT*	*E: BOT*	*F: Corridor management*
Public sector risks (land acquisition and relocation risks always carried)	Design Construction Traffic and revenue levels	Planning Traffic and revenue levels	Revenue Macro Some regulatory	*Force majeure* Some regulatory	Planning Macro Some regulatory	Planning *Force majeure* Macro Some regulatory
Typical contract size ($)	Small	Medium/large US$50–US$800 million	Small/medium	Medium/large	Very large c. US$100 million to US$1 billion	Medium/large c. US$90–US$300 million
Minimum size concessionaire required	Small/local construction firm	Small/local construction firm	Construction firm with management skills	Larger construction firm with management skills	Consortium including major construction firms	Consortium often with major construction firms
Typical duration	2–10 years	Defined construction period	2–10 years	10–20 years	c. 30 years	c. 30 years

Note: For more details and some differences, see also ADB (1999).
Source: Authors.

Box 6.2. *Increasing Private Participation in Roads: Argentina's Experience*

The general privatization strategy was to unbundle financially viable roads into build-operate-transfer concessions awarded through competitive bidding. Most of the traffic is concentrated near major city nodes, such as Buenos Aires and Rosario and Córdoba to a lesser extent. The national concession program has so far focused on the multilane roads and freeways serving these cities, along with other intercity and major city access roads. It applies now to almost 9,500 kilometers of 38,000 kilometers of national roads. The concession program was complemented by an auction of management contracts (generally for five years) for rehabilitation and maintenance, now covering about 12,000 kilometers of national roads divided into 400 sections and auctioned out into 61 contracts. Also, nontoll concession contracts cover about 1,900 kilometers of national roads (six corridors) and allow the government to rely on a private financing of the initial rehabilitation in exchange for a commitment to future disbursements of monthly subsidies during the 10-year terms of the concessions. A more recent program called "km/month" covers basic maintenance and service contracts for 4,100 kilometers of less traveled roads. Overall, about 70 percent of the national road network is de facto under private operation.

tolls, while the public sector retains legal ownership and regulatory oversight of the concession contract. In most projects, design responsibility is shared, with the public sector taking the lead in corridor identification and preliminary design, leaving specific details to the private sector, subject to government approval. In practice, the government often ends up sharing some of the demand risks through the payment of subsidies. Typical BOT concessions are 20 to 30 years in length, whereas maintenance concessions tend to be shorter, typically 5 to 15 years. They differ in length because of the different financial requirements. The duration of the concession may either be set in advance by the government or be part of the decision criteria in selecting the concessionaire.

Overall, BOT concessions are most likely to be successful under the following conditions:

- Projects minimizing costs in existing high-traffic corridors, for example, projects with missing links such as river crossings, because they minimize land costs; inter-urban projects with low implementation costs; and urban area projects at grade or elevated, because they keep construction costs low
- Projects in countries where there is a tradition of paying public tolls, or at least where the willingness to pay the proposed toll level has been carefully assessed
- Projects in which tolls are set at, or close to, the revenue-maximizing tariff and toll escalation formulas are invoked

- Projects that have an existing income stream from which to draw revenues from day one, perhaps even during the construction period.

Risk Allocation Options

Choosing among the options for private participation shown in table 6.3 depends on the particular needs of a country and the nature of risk sharing between the public and private sectors. Risk allocation is a complex and difficult process, and for all practical purposes, it is a negotiated process (for a much more detailed analysis see Irwin and others 1997). Unfortunately, these initial negotiations seldom involve the future regulators, even when their outcome is critically important to regulatory decisions. This is why one of the first tasks a new regulator has to address in its new position is to understand the distribution of risks to which each party is committed through the contract, because in many renegotiations or regulatory disputes, the responsibility will be based on the assignment spelled out in the contract.

The rule of thumb is that private road infrastructure projects work best when project risks and responsibilities are assigned to the party that can best bear them. The private sector generally is better at managing commercial risks and responsibilities, such as those associated with construction, operation, and financing. In contrast, toll roads may also depend on public participation in areas such as acquisition of right-of-way, political risk, and in some cases, traffic and revenue risk. Successful projects have been characterized by a broad level of risk sharing between the public and private sectors. Privately supported toll road projects work best when experienced, well-capitalized firms have some discretion over design and confidence in toll policy to accept construction and some degree of traffic risk. The government assumes the risks that it controls and considers giving financial support or guarantees if traffic levels in the early years are insufficient.

In practice, this theory of risk allocation is often not applied. Part of the reason is that risk levels and types tend to change. The 1998 Asian crisis sufficiently increased risk levels worldwide, increasing the cost of capital to unbearable levels for many potential investors. Governments can also be subject to a fear-greed cycle in which they become afraid of program failure, and thus offer increasingly better terms. Prospective concessionaires may worry that they will be left out and end up making unrealistically optimistic bids. Subsequently, the element of greed takes over and governments may fail to live up to commitments, and the private sector seeks ways to privatize gains and socialize the project risks.

The main risks facing toll road projects are preconstruction activity, construction, traffic and revenue, currency, *force majeure*, tort liability, political risk, and financial risk. The privatization teams must address these risks in a satisfactory manner before debt and equity investors will commit to project funding. The standard risks that contracts identify are preconstruction, construction, traffic and revenue, financial, regulatory, and political. In addition, contracts commonly address *force majeure* and legal liability because they have proven to be serious sources of cost overruns in the sector.

PRECONSTRUCTION RISKS. Many projects are delayed because of the difficulties of acquiring right-of-way or environmental clearance that both the governments and the operators underestimated. The most relevant effect is cost overrun during project development. In general, the public sector often ends up taking on the responsibility for most of these risks, because acquiring the right-of-way, paying for it, and contributing this asset to the project are often easier for the public sector. Problems often arise when the government is not providing the road itself. If a private sector partner is undertaking construction, the delineation of responsibility and phasing of development by the different parties is particularly important.

Regulators end up having to address this kind of risk, as seen in the following typical experience. A new segment of the Don Muang Tollway in Bangkok, Thailand, will connect the airport with another toll road. In 1989, the Department of Highways gave the Don Muang Tollway Public Company Limited (led by a German firm) a 25-year concession to build the US$407 million, 15.4-kilometer initial segment of the project. One clause in the concession agreement specified that the government would remove flyovers on a parallel road that competes with the toll road and would then construct new flyovers to allow radial movement. The government did not deliver for more than two years, however. In addition, it blocked toll rate increases until the completion of the new flyovers. As a result, toll revenues were almost 30 percent lower than had been forecast for the period. The sponsor ended up close to bankruptcy, which forced the government to provide significant compensation in exchange for a 40 percent stake in the company, thereby helping refinance the loans (ADB 1999). One way to reduce transaction costs would have been to come up with a clearer contractual commitment for the government to take on that risk, and possibly to have it put a guarantee fund together to establish the credibility of its commitment, in the same way that governments ask concessionaires to fund commitments through guarantee funds. The general principle is the same: credible, rule-driven decisions are always easier for the regulator to implement.

Construction Risks. A common cause of cost overrun stems from design changes and unforeseen weather conditions during the construction phase. For instance, between the time that a concession is signed and when the concessionaire takes over the business, a hurricane can significantly increase construction costs. Who should pay for the consequences of the hurricane? The private sector typically bears primary responsibility for such risks and may attempt to cover some of them through insurance. The public sector may assume responsibility for risks under its control, however, such as completing complementary facilities (connecting roads or interchanges) or allowing cost increases associated with major design changes. Commonly, governments also at least share costs for projects that face major construction uncertainties, such as toll roads through mountains.

Most cases, though, use fixed-price construction contracts, with some provision for severe disruptions. For example, in Brazil, a financial equilibrium clause enables contractors to renegotiate contract terms if major design changes are required. When massive cost overruns occur, contract renegotiation may be required in exchange for sponsors and creditors providing additional financing. This occurred in the Guangzhou-Shenzen project in China, where the private sponsors made an additional US$700 million equity investment in exchange for an increase in the profit sharing agreement during the first 10 years of operation.

Note that that the use of fixed construction prices in the contract is consistent with the idea of facilitating the work of regulators, but it also illustrates the costs and risks involved with accepting rules too readily. Concession units being staffed with members of the public roads department is not uncommon. In some countries, this staff has an established contact with many of the local construction companies through procurement and maintenance contracts for the public roads, and the bidding rules for these contracts are not as competitive as they should be, resulting in construction prices that are not consistent with best practice. Thus, the risk is that unit prices built into concession contracts are based on the wrong prices (in the best of cases), or that they reflect collusion between the concession unit and the concessionaires (in the worst case scenario). In other words, regulators should not always take construction unit prices for granted when they have the option to review them.

Traffic and Revenue Risks. Demand uncertainty continues to be a major problem at the conception stage and ends up haunting many, if not most, projects. Traffic and toll levels may not be sufficient to cover all costs, including construction, operation, and maintenance. An approximate rule of thumb is that 10,000 to 15,000 vehicles per day (vpd) are needed to fully cover operating and capital costs. Coverage of operating costs alone generally requires traffic in excess of 3,500 vpd. Recovery of toll collection costs requires approximately 1,500 vpd (Fayard 1993). The handling of traffic and revenue risks ranges from full private sector assumption to government-provided traffic and revenue guarantees. The policy issues involved with managing these risks are major strategic choices that this chapter later discusses in detail, and they vary tremendously depending on the time and location of the project.

A regulator's main concern in this context is to make sure that it has access to the demand studies conducted in preparation for the tolling of the road network. As explained earlier, forecasting demand is a challenging task that privatization teams often underestimate. Overoptimism is common for privatization teams that focus on convincing private operators of the value of their business and for potential operators that want to make a deal and are convinced that they can renegotiate almost anything once they have taken over the business. To be somewhat credible, much more so than in most other transport studies, one has to combine analysis of the willingness to pay for a toll road with a study of ability to pay to fairly assess the traffic and revenue risks. In many toll road renegotiations, the regulator's main concern is to avoid boycotts of the road by users who are unwilling or unable to pay for the toll. The solution is often to cap the toll and adjust the duration of the contract, but the adjustment often entails significant transaction and political costs that most regulators wish to avoid.

CURRENCY RISKS. The impact of exchange rate fluctuations on the value of the business drives the main currency risk. In addition, the toll concession can be subject to a convertibility risk that refers to the possibility that the operator may not be allowed to exchange local currency for foreign currency. These are major issues for toll roads that are financed with foreign capital, because revenues are commonly in local currency and adjustments for inflation and exchange rates may lag or encounter political opposition. Projects can reduce this risk by tapping domestic capital markets when possible. Most projects attempt to mitigate exchange risk by including provisions for indexing to inflation, although in practice, the magnitude of exchange rate volatility has made such requirements difficult to enforce.

Peru, for instance, addresses and shares this risk in concession contracts in the following way. To begin with, the initial basic toll unit is expressed in dollars. This tariff is adjusted every six months in line with the consumer price index using a devaluation index that the National Statistics Office publishes. The devaluation adjustment only kicks in when the devaluation rate is higher than inflation. The toll is adjusted by 50 percent of the difference between devaluation and inflation. A general formula would look like this:

$$P_{tMN} = P_{t-1\,MN} * (1 + CPI_{t-1} * [1 + \beta * (DEV_{t-1} - CPI_{t-1})]$$
$$\text{if } DEV_{t-1} > CPI_{t-1}$$

where $P_{t-1\,MN}$ is the toll base adjusted in national currency for the period $t-1$; β is the factor by which the difference between devaluation and inflation can be passed on through tolls, which essentially is the variable over which a negotiation takes place between the government and the concessionaire; CPI_{t-1} is the consumer price index in the period $t-1$; and DEV_{t-1} is the devaluation in the previous period.

Having an explicit formula like this is always a blessing for regulators and is now becoming standard in concession contracts, so that when an explicit rule is not available, regulators only have to check compliance rather than arbitrate a negotiation between the government and the concessionaire.

FINANCIAL RISKS. Financial risk is the risk that project cash flows might be insufficient to cover debt service and then pay an adequate return on sponsor equity. Financing constraints, especially the lack of long-term debt capital, significantly hinder toll road development. Since the advent of financial crises in emerging markets, few projects have been able to generate returns on investment that are sufficient to attract private capital. Required debt ratios have fallen from 70 percent to 40–50 percent, with costs of capital rising to 20 percent or more. This suggests that until macroeconomic risk premiums decline and traffic growth is more established, only the highest-density projects will be undertaken without substantial government support. The financial crises will force many programs to slow down and force debt restructuring of many existing concessions. The promotion of more secure financing structures is needed to reduce the risk of potential bailouts.

Because toll roads are long-lived investments with high start-up costs, countries with local capital markets that can provide long-term financing have many advantages in supporting toll road concessions. Of particular importance is the available maturity of domestic finance. In many countries, new toll concessions have been unable to obtain financing for longer than five to six years, which creates a major refinancing risk that either renders the project nonviable or requires government guarantees of such a rollover.

In theory, financial risk is best borne by the private sector, but in toll road projects substantial government risk sharing is likely, either through revenue or debt guarantees, or through participation by state or multilateral development institutions. Cash grants or other financial contributions also may be available, which serve to improve the project's rate of return on private finance.

REGULATORY RISKS. Regulatory risk stems from the weak implementation of regulatory commitments built into the contracts and the laws or other legal instruments that are relevant to the value of the transaction as it was originally assessed. Essentially, the question is whether the regulator will exercise its authority and responsibilities over prices, public obligations, competition rules, and similar rules that the contracts specify, and whether that will influence the value of the business. This risk is more common than it appears, and pressures on regulators are a major source of concern that investors incorporate into their required rate of return. In 1999, a major factor in the restructuring of Mexico's toll road program was the pressure on regulators to cut tolls. In Thailand, a similar concern resulted in the government's decision to cut a toll level by 50 percent of what it had committed to in a BOT contract. The outcome was that the government ended up taking over the toll road.

The solution is to try to make sure that regulators have rules to follow and that they are independent enough to be able to enforce them. First, the rules must cover the possibility of adapting the contract terms during the concessionaire's tenure. Toll road concessions tend to be long, and the legal environment in reforming countries tends to change during that period. For instance, environmental and safety concerns are increasing in many countries. New laws are introduced during the term of many toll roads. The rules that allocate the financial consequences of these changes among government, users, and operators are critical, yet often forgotten.

Even if regulatory rules are clear, they are only as effective as the regulator. The best designed regulatory contract is useless if the regulator is not independent or fair, which has been a major source of concern in Brazil. For example, in a concession between the cities of Rio de Janeiro and Teresopolis, illegal access and egress has been estimated at 3,000 vehicles per day. The mayor of Mage, a small town along the route, has championed this leakage, because he believes that his citizens should not have to pay what are perceived to be very high charges for local access users. Regulators have not been able to enforce the contractual commitments made to the operator.

POLITICAL RISKS. Political risk concerns government actions that affect the ability to generate earnings. These could include actions that terminate the concession, the imposition of taxes or regulations that severely reduce the value to investors, restrictions on the ability to collect or raise tolls as specified in the concession agreement, and the preclusion of contract disputes to be resolved reasonably. Governments generally agree to compensate investors for political risks, although in practice, governments may cite justifications for their actions to delay or prevent such payments. Thus, private investors generally assume the risks that are associated with dispute resolution and the ability to obtain compensation if the government should violate the concession agreement. The issue of meeting financial obligations while disputes are resolved may be achieved by requiring debt service reserves, escrow, or standby financing.

In Brazil local political interference has affected several toll road projects. A state-level concession in Parana is the most significant example to date. In this case, new tolls were introduced during peak harvest season and the governor forced the concessionaire to charge only 50 percent of the original tariff. The case is now nearing a decision in court, but all the other concessions are paying close attention. If the original toll contract structure is not fully upheld, accomplishing refinancing and attracting capital on favorable terms for a second wave of concessions will be more difficult. Investment bankers have cautioned that if these court issues regarding toll revisions in Parana are not resolved, an additional 200 basis points could be required for those projects in regions with particularly populist governors or mayors. In total, including the costs of the spillover effect of the Asian crisis in the rest of the world, the costs of debt rose from approximately 11 percent in late 1998 to 16–17 percent in early 1999. At this cost of debt levels, most projects are not viable at their planned toll levels.

The government's credibility to uphold contractual obligations and its willingness and ability to provide compensation for political risks are key issues for private investors in toll roads. Issues with delays or denials of toll increases have made many prospective parties wary of entering into new projects. This is especially true for foreign capital, which is perceived as especially vulnerable to political risks. Some of the more risky emerging markets may require support from multilateral or bilateral financial institutions to reduce this risk exposure. In addition, political risk insurance may help manage issues of inconvertibility, transfer, and confiscation. Box 6.3 shows how a regulator might put together all these risks into a single quantitative indicator.

OTHER RISKS. *Force majeure* refers to risks that are beyond the control of both public and private partners, such as floods or earthquakes, that impair the project's ability to earn revenues. While some private insurance is becoming available for catastrophic risks, the public sector generally is faced with the need to restructure the project should such disasters occur. This

Box 6.3. *How Should a Regulator Consider Risks?*

Risk factors can be pulled together in the concept of cost of capital, which represents the required rate of return that all investors blended together might expect on a project. For most regulatory decisions, a regulator will have to assess the impact of its decisions on the cost of capital through its impact on each one of the risk levels. Algebraically, we can simplify and write this as follows:

Cost of capital = (required rate of return on debt) x (percentage of debt in the project)
+ (required rate of return on equity) x (percentage of equity in the project)

Because interest expense typically is tax deductible, we can calculate the cost of capital either on a before-tax or an after-tax basis. It is important to understand that the tax rate that is relevant is the one that applies to project sponsors.

The required rate of return on debt. The required rate of return on debt (that is, the borrowing cost) includes a number of risk factors, each of which commands a premium that must be paid to investors in order for them to bear that particular risk:

Required rate of return on debt =
risk-free borrowing rate for specified time horizon +
premium for country/financial risk +
premium for currency risk +
premium for project or sector risk (including construction) + premium for regulatory risk

The required rate of return on equity. Similarly, the required rate of return on equity investment can be seen as being equal to a risk-free rate plus a premium for the higher risk faced by equity relative to debt, as well as all four risk factors above. The equity risk premium is a function of how risky a specific sectoral investment is relative to equity markets overall. (This adjustment factor is known as beta and has an average value of 0.6–0.8 for toll roads.) Thus,

Required rate of return on equity =

risk-free borrowing rate for specified time horizon +
equity risk premium (adjusted by project beta) +
premium for country/financial risk +
premium for currency risk +
premium for project or sector risk (including construction) +
premium for regulatory risk

While in many cases the risk premiums required would be similar for debt and equity, this will not always be the case. For example, regulatory lags in approving pricing decisions may have a greater effect on equity holders because creditors have a prior claim.

may take the form of extending the concession term or providing additional financial support. The rule is that contracts should state remedies in the event of *force majeure* risks, for example, cash compensation or an extension of the concession term equal to the length of the disturbance. Finally, tort liability refers to liability for legal awards as a result of accidents or negligence on the toll road. This responsibility is borne by the private sector and is typically covered through private insurance. Governments, however, should make sure that such coverage is adequate and that the insuring party is financially sound.

Regulatory Options for Mitigating Risk

At the start of the concessioning process, the government has two main reasons to commit to supporting toll road projects at the beginning of a project: (a) to offset the financial or exchange risks by reducing capital expenditures, or to improve revenues to the extent necessary for a project to cover debt service and provide a reasonable equity return; and (b) to offset the demand and traffic risk and protect investors, especially lenders, from the risk that actual cash flows will fall below expected cash flows and thus be inadequate to cover debt service. When unexpected events occur and renegotiation of a contract arises, these two are often the main problems that a regulator must address. The name of the game is to come up with a mix of government actions that ensures that an acceptable financial return can be generated, such as some redesign of the financing schemes to include guarantees, as well as redoing the project design, including its duration.

THE VARIOUS INSTRUMENTS AVAILABLE TO A REGULATOR. If public financial support is appropriate, one can use a variety of mechanisms to support private toll financing. These instruments range from revenue enhancements to equity guarantees as follows

- *Equity guarantees:* These provide a concessionaire with the option to be bought out by the government at a price that guarantees a minimum return on equity. Although the liability is contingent, the government effectively assumes project risk and reduces the corresponding private sector incentives.
- *Debt guarantees*: These guarantee that the government will pay any shortfall related to principal and interest payments. The government may also guarantee any scheduled refinancing. This creates significant government exposure and reduces private sector incentives, although it may decrease the cost or increase the amount of debt available to the project.
- *Exchange rate guarantees*: These are when the government agrees to compensate the concessionaire for increases in financing costs due to exchange rate effects on foreign financing. Exchange rate guarantees expose the government to significant risk and increase the incentive to use foreign capital.

- *Grants/subsidies*: These are contrary to equity and debt guarantees that create contingent liabilities for the government. Alternatively, governments can furnish grants or subordinated loans at project inception, buying down the size of the project that needs private finance. (In Chile, the size of the government grant was one of the criteria used in awarding the south access toll road concession.) Alternatively, explicit subsidies can be given as part of the renegotiation process. In Argentina, this subsidy took the form of the forgiveness of accumulated payments due to the government for the right to operate the concession. In general, these grants or subsidies have no provision for repayment.
- *Subordinated loans*: These can fill a gap in the financing structure between senior debt and equity. From the government's perspective, they also have the attractive feature that they can be repaid with a return if the road is successful. Subordinated loans improve feasibility by increasing the debt service coverage ratio on senior debt and by reducing the need for private equity, which requires a higher return. Because subordinated debt does eventually require repayment, however, it does not improve project feasibility to the same degree as a similarly sized grant. Another alternative would be for the government to contribute financing that has characteristics of both debt and equity. One such instrument would be a so-called reverse convertible contribution that would remain as equity unless the project was successful, at which point it would convert to debt for repayment.

As an alternative to these instruments, the regulator could rely on "playing" with the design of the contract. This involves considering changing the time profile of toll revenue as well as the toll levels and types, or adjusting the investment specification and other service obligations or the contract duration as follows:

- *Minimum traffic and revenue guarantees:*[10] These are a relatively common form of support for toll roads in which the government compensates the concessionaire if traffic or revenue falls below a minimum threshold. Typically the threshold is set 10 to 30 percent below

10. Note that some countries, such as Chile, jointly introduce minimum imcome guarantees to protect the operator with a revenue sharing scheme that allows the government a 30 to 50 percent share of extra profits (in other words, revenue that generates a return in excess of 15 percent) when traffic is consistently above what was forecast.

the expected volume, and relying on a revenue guarantee is generally more desirable if the goal is to facilitate the operator's access to the financial market. This trigger reduces government exposure while providing sufficient revenue coverage to support the debt component of the capital structure. In addition, traffic and revenue guarantees help retain financial incentives in the project, unless conditions deteriorate well below what was forecast. If the government shares downside risk with the private sector through guarantees, it should also consider seeking instruments that allow profit on the upside. One way to do this is with a revenue-sharing arrangement in which the government receives a portion of revenues above a maximum traffic threshold.

- *Shadow tolls:* These are a way to provide subsidies in which the government contributes a specific payment per vehicle to the concessionaire. In effect, they are an ongoing revenue stream from the government in lieu of an up-front grant or loan. Because they are paid over time, they may be less of a burden on the public budget. The drawback of shadow tolls is that they may not provide investors with much protection from revenue risks. That is, shadow toll payments are highest when traffic volumes are large. As a result, government payments may be inadequate to protect investors when traffic is low and may be unnecessarily high when traffic volumes are high. In addition, the payment of shadow tolls over time creates a credit risk for concessionaires. One can reduce these inefficiencies in a number of ways, such as by implementing a declining payment schedule as volumes increase or a maximum traffic level beyond which shadow tolls are not paid. Because they tend to top off private revenues, shadow tolls may be particularly valuable as support to low-volume roads that require upgrading or rehabilitation rather than new construction.
- *Concession extensions and revenue enhancements:* These provide financial support that involves limited public sector risk, but they do little to support or enhance private financing. First, a government can extend the concession term if revenues fall below a certain amount. Second, a government can restrict competition or allow the concessionaire to develop ancillary services.
- *Changes in contractual obligations:* These allow the redesign of contractual obligations. Slower or less investment and fewer service obligations are ways to cut costs and transform a nonviable road into a viable one.

Choosing Among these Instruments. In general, the most advantageous types of support for the concessionaire are those that provide early funding streams (when toll road revenues are low or nonexistent during the construction period) and those that give guarantees for unexpected problems (for example, exchange rate guarantees). This is true at the time the contract is initially signed as well as whenever the regulator is asked to renegotiate to restore financial viability to a project that has lost its viability. The least significant are those that themselves are unpredictable, such as additional rights for development around the road. One can use these various mechanisms of government support in combination when a project is not feasible on its own and where revenue risk is substantial. In such cases, grant plus minimum revenue guarantees may be sufficient to induce private participation. Governments should avoid broad guarantees that reduce lenders' scrutiny and due diligence. In many cases, the availability of these guarantees have induced lenders to provide funds based on guarantees and sponsor strength rather than on underlying project risks and revenues.

When assessing the value of these adjustments, regulators must recognize that the value of government support also depends on the credibility and credit risk of the government itself. Investors may be inclined to discount the value of various support mechanisms that have not been upheld in the past, or which are tendered for long periods. Governments also need to improve the management of their contingent liabilities in order to maintain their fiscal credibility, and thereby reduce macroeconomic risks that directly affect toll roads through traffic volumes and financing costs. However, governments are sometimes tempted to increase support far above expected levels when the sponsors are well-connected politically, have better advisers, or threaten to withdraw at the last minute. To prevent this, the government should be well prepared with the specification and design of its part in support of that preparation. The upshot is that determining if a project requires government support and how such support should be structured requires a detailed analysis of project costs, revenues, and risk, as well as an understanding of what debt and equity investors require. Most regulators have ignored the importance of this information and have not been able to appropriately monitor or arbitrate disputes as a consequence. Before bidding a concession, governments should be aware of the project's critical elements, including environmental issues, traffic and revenue potential, preliminary design and costs, permit requirements, and the views of potential investors. Governments can improve the likelihood of having successful projects by undertaking studies of these issues and by working with experienced advisers. Box 6.4 tells how Peru effectively prepared its toll road program. Unfortunately, a lack of political commitment to the program is still delaying its implementation. The regulators will, however, have all the required information once the program is implemented, thanks to effective preparation.

Box 6.4. *Preparing for a Toll Road Program: A Lesson from Peru*

Faced with rapidly growing motorization, in 1997 Peru decided to launch new initiatives in road transport and to transform its public tolled highway network into a wider private tolled network. A special committee quickly began the process of selecting consulting firms to undertake studies of the existing national road facilities, as well as demand and detailed engineering studies for an expanded system of national toll roads to be offered through a system of concessions. Using the existing toll network as a base for expansion, the engineering and very preliminary demand studies led the special committee to designate 12 prospective concessions, totaling 6,750 kilometers. Estimated cost for the total network of improvement and expansion is US$1.1 billion. Most of the proposed concessions incorporate segments of the existing toll road system. Each new proposal develops a plan for upgrades and expansion, and then grafts an additional new segment on to this base road.

The result is a set of concessions for which prospective traffic volumes will vary enormously over the different road segments. This creates concessions that, by design, have included cross-subsidies of low-density segments with high-density ones. The essential assumptions of this preliminary study included a traffic growth rate of 3 to 5 percent per year; periodic maintenance costs per kilometer every five years between US$10,400 and US$14,500, depending on the road; rehabilitation costs around US$100,000/kilometer and reconstruction costs of US$350,000/kilometer. Tolls would be set at US$2/100 kilometers and would automatically be adjusted for inflation and exchange rates. (The precise mechanism for dealing with the interaction of inflation and exchange rates remains to be settled.)

These assumptions allowed an estimate of the net present value of toll revenues (net of operating and maintenance requirements). Subtracting the estimated net present value of net toll revenue from the estimated net present value of the investment (excluding land costs) yields the estimated new present value of each road project. Only 3 of the 11 projects have positive net present values at a 15 percent real discount rate in dollars. Those three proposed concessions incorporate sizable amounts of the existing toll network, and as such, face relatively low expenditures on land and improvements. Notably, even on the perimeter of Lima, high investment costs overwhelm higher traffic density.

The preliminary studies indicated that low traffic volumes and large required investments would not allow concessions to be let on the basis of financial payments to the government. The result was the development of a negative concession plan. Concessions would be bid on the basis of the lowest amount of investment the central government would make, and they would run 25 to 30 years, with subsequent transfer of the roads to the government. The government's contribution would not be considered part of the equity in the concession. The government would delegate the responsibility for the enforcement of the contract to a transport regulatory agency that would resolve disputes or pass them on to the judicial system.

Contract Design from a Regulatory Viewpoint

The concession agreement is the principal contract governing a private toll road project. One can design it in many ways. (See also Fishbein and Babbar 1996.) In some countries the government provides many of the details in the information sets provided to the bidders, and the bids are for specific proposals. In other countries the government asks the bidders to make many of the suggestions to implement the road. Whatever the sequence, the following is a minimum list that the overall contract package needs to cover to allow the regulator to referee in cases of conflicts between users and the concessionaire or the government and the concessionaire:

- *A definition of the legal context.* Toll road projects, whether wholly private or mixed in character, require a clear legal context defined by well-drafted laws and regulations regarding concessions. The policy framework should address the types of roads targeted for tolling, the types of organizational structures allowed, and which government entities are responsible for overseeing the program. Because many different forms of toll road development exist, the legislation may be general in character, enabling different types of private participation. Why should a regulator care? Because these laws must clearly identify the respective rights and obligations of the private and public sectors, which is a crucial element of the settlement of any dispute between the concession agency and the concessionaires. Also fundamental is that these rights and obligations are seen as valid, binding, and enforceable through a legal process that is fair, timely, and not overly costly. In addition, the regulator needs to be informed how the toll road program is integrated with national, regional, and local transport policies and is enabled by a concession law. For a toll road program to be effective it must be coordinated with broader transport and road policies. The entire process should be designed to be competitive, transparent, and based on reasonable evaluation criteria.
- *The administrative background.* As with any type of contract, the regulator must be able to refer to a set of definitions for all the key concepts the contract covers . This includes such items as the definition of the concession area, the zone of added services, maintenance, what constitutes *force majeure,* what constitutes basic or special services, the key monetary and technical units, the standards to be used, and the key players involved in the sector. From a regulatory

viewpoint, of particular importance is the contract specification of those events that would constitute default on the part of each party, including remedies and the procedures for obtaining compensation. Finally, the administrative requirement may also have to provide a definition of what constitutes the basic documents that give all the required information to all parties involved. The minimum set includes the explanation of the administrative, technical, and financial requirements. Increasingly, countries are also including in this definition any ulterior clarification to be issued as a result of mistakes potential bidders identify when reviewing the documents. Taken together, these documents provide the basis of the information to be used by the regulator.

- *Estimate of the costs of the project.* The regulator needs to get an idea of the value of the task at stake. In some cases, this results in the unit costs and the maximum cost of the project as estimated by independent engineers, which the bidding documents should also specify to provide a benchmark. Often the government will have several independent studies that include both demand and cost studies.
- *The asset valuation rules.* The government should be interested how the assets are evaluated for fiscal reasons as well as for regulatory reasons. Indeed, the value of business will be at the core of many regulatory decisions involving the toll level or the duration of the contract.
- *The economic content of the technical documents.* The technical documents must cover at least a few items that the economic regulator needs to sort out the financial and economic consequences of the operator's actions, whether imposed by the bidding documents or proposed as part of the bid. The main aspects are the investment and maintenance plan and timetable and the toll system description (including technology and location). They should also cover information on weights allowed for each type of vehicle, which is relevant for the calculation of the maintenance costs and related toll levels. The documents should also clearly define the rules of the game for the evaluation of these technical bids to allow the regulator to settle any related dispute.
- *The various types of guarantees and warranties.* This section frequently includes requirements regarding insurance, performance bonds, minimum equity contributions, and corporate structure. They may apply to all stages of the process (offer, construction, and operation) and generally cover specific amounts for the various stages and apply to both the concessionaire and the government. For the government, they

may include commitments regarding approvals and right-of-way permits, expropriations, and so on. These sections provide one way of telling the regulator how much is at stake in the decisions regarding compliance with obligations on all parties to the contract. In principle, the guarantees should have an economic meaning in the sense that the amounts involved should somehow be related to the risks of noncompliance, but in practice, they are seldom related. They tend to be somewhat arbitrary amounts, negotiated to be large enough to induce private participation or financing.

- *The identification of the various types of risks and their distribution between the parties.* This section typically covers each party's specific responsibilities for funding, acquiring, and preparing the right-of-way, including risks of delay or cost overruns. It also includes responsibilities for developing and constructing the project, including environmental compliance, permits, and designs. The agreement should address the risk borne by each party in the event of unplanned delays, cost overruns, and so on. In addition, the agreement should address the possibility that financing will not be raised. The contract also should specify any rights or responsibilities of the concessionaire to modify or expand the road in the future beyond the requirements of the initial concession. The agreement should specify the conditions under which profits or revenues are shared with the government. For example, if using a maximum traffic or revenue ceiling, the agreement should state the maximum traffic or revenue threshold for each year of the concession, the revenue sharing formula, and the procedure for calculating and transferring the payment to the government. If using incentive provisions, the agreement should specify the events that would trigger the incentive payment and the size and timing of such payments.
- *Concession rights and obligations.* These should include an explicit definition of the concessionaire's exclusive right to design, build, finance, and operate the project during the concession period, which will provide the regulator with basic benchmarks to assess compliance with commitments. The contract should include the service obligations (for example, farmers can use some portion of the road for free) and related compensations to which the operator is entitled, the conditions under which the concession may be extended or amended, any payments required either by the concessionaire or by the government, and specifications as to who holds the legal title and how any transfer will occur. The concession contract should define the responsibilities of each party

for operations, including toll collection, maintenance, enforcement and safety, auxiliary services, and administration. The contract also should make explicit (and ideally, formula-driven) any mechanisms the government commits to support the project, including magnitude, timing of payments, duration of support, and conditions under which support is phased out or withdrawn. It should address specific facilities such as connecting roads or interchanges that the government or concessionaire is committed to provide, including dates and remedies in cases of delays or nonperformance. This section also should define the recourse of the concessionaire should the government not honor its financial commitments under the agreement.

- *The penalty rules.* In addition to relying on the threat of cashing in deposits for guarantees, regulators need to have access to a clear set of fines that relate the penalty for noncompliance on more operational matters to the damage resulting from the noncompliance. Here, the practice seems to be to set predefined amounts for specific types of violations to minimize the arbitrariness of regulatory decisions. The concession contract for Road 5 from Santiago to Talca, Chile, identifies and defines 81 types of violations and specifies the amounts involved and the application criteria (such as every day, every time, and so forth). To the extent possible, and to make regulation easier, established performance standards should relate to the penalties for noncompliance.
- *The regulatory regime.* The contract must specify the regulatory approach and enforcement mechanism. If using rate of return regulation, the agreement must specify the basis for the regulation, the maximum rate of return allowed, and the calculations required to monitor the concession performance. If using toll rate regulation, the agreement should specify the maximum toll by vehicle type, the index used to adjust toll rates, and the time period for toll rate adjustments. Some degree of creativity is allowed here. Peru, for instance, adjusts the standard formula to include a premium for improvements in safety over the targets the contract spells out. The contract also should include the specific procedure for calculating and revising the toll schedule (specific pricing rules are discussed later).
- *The information the operator will be required to provide to the regulator.* The contract should specify the type and timing of information to be provided to the government to monitor the agreement. The contract should also specify the conditions under which the regulator can ask for additional information not covered by the contract. Typically, the regulator

will ask the operator to provide monthly data reports on hourly, daily, and monthly vehicle flows, classified by vehicle type, as well as monthly data reports on congestion, accidents, and changes in regular traffic patterns. In addition, quarterly reports on auxiliary services will provide sufficient information on any related service obligation the contract imposes. Every six months, the regulator should expect reports on maintenance costs, actions taken, and total and unit costs, as well as a report on paving progress if the contract specifies this.

- *The acceptance conditions.* The contract should specify the conditions under which the government will accept the completed facility and approve the start of operation. This is particularly important when tolling is scheduled to begin before the project is completed. In Brazil this approach provided a way to generate early revenue, while allowing the public to see the improved road before having to pay for it through tolls.
- *Limitations on competing facilities.* The contract should specify the corridor, if any, under which the government is restricted from constructing, expanding, or granting concessions for competing roads or other facilities. As mentioned earlier, the existence of free parallel roads is a matter of concern for many operators, and regulators may have to arbitrate challenges by governments to operate almost parallel routes.
- *Rights to access third-party operated facilities.* The contract should spell out any specific rights of the concessionaire to access land or roads owned by third-party activities as part of the concession, including how to pay for this access. In most conflicting events, the regulator will be called to assess the access pricing rule demanded by the owner of the facility to be shared.
- *Assignment and termination of the concession.* The regulator also needs to have clear instructions on the terms and conditions under which the concession may be transferred to a party other than the original concessionaire, including the specific conditions under which the concessionaire or the government can cancel the concession and the consequences of termination, including penalties and replacement.
- *The renegotiation rule.* Renegotiation happens. It is actually quite common and the contracts should be clear and try to have preestablished rules to avoid the conflictive situations that were frequently observed in relation to infrastructure contacts in the early 1990s. More recent contracts carefully spell out these rules in Latin America. Chile's example, which box 6.5 discusses, may be the best so far.

***Box 6.5.** Rule-Based Renegotiations: Lessons from Chile*

To provide flexibility without compromising the concessionaire's interests, the Chilean contracts include detailed procedures to constrain and financially assess government requests for additional work. The government can demand additional work for up to a maximum of 20 percent of the initial official cost estimate of the project, up to two years before the concession ends. During the construction stage, the government can only demand additional work for up to 5 percent of the official cost estimate, and new investment at that stage is valued according to a unitary pricing schedule contained in the tendering documents. Bidders implicitly accept these unitary costs when they participate in the franchising process.

The Ministry of Public Works and the concessionaire must agree on the valuation of new investments required during the operational phase. If they do not agree, differences must be settled based on technical reports that consultants from each party produce. The compensation can be through increased tolls, increased duration of the concession, or direct payments by the state.

To avoid conflicts, the most recent concessions place explicit restrictions on the compensation mechanism. For example, in the Río Bueno–Puerto Montt concession, tariff increases during the life of the contract cannot exceed 25 percent, and the increase in concession duration cannot exceed 120 months. Furthermore, the contract includes an explicit formula to calculate the required compensation. This is given by

$$\sum_{i=k+1}^{N+S} \frac{Y_t - T_t}{(1+r)^{i-k}} = I_k + \sum_{i=k+1}^{N+S} \frac{C_t}{(1+r)^{i-k}}$$

where I_k = additional investment in period k, N = initial duration of concession, S = extension of contract, and Y_t = additional income due to increase in tariffs, where

$$Y_t = \begin{cases} \rho_t P_t Q_t + G_t & t = k+1, \ldots, N \\ (1+\rho_t) P_t Q_t + G_t & t = N+1, \ldots, N+S \end{cases}$$

and P_t = tolls prior to compensation, Q_t = projected traffic levels for new investment at initial toll levels, ρ_t = percentage increase in tolls,G_t = direct payments by state, C_t = operational and maintenance costs associated with new investment, T_t = taxes due on additional toll income, and r = discount rate.

The additional operational and maintenance costs, the projected traffic levels, and the discount rate must be based on an expert's report. If disagreements arise over these parameters, the Conciliatory Commission must convene. The tender documents are usually more explicit on how to estimate the discount rate, however, and they place an upper limit on the risk premium that the concessionaire can receive.

To avoid imposing additional traffic risks on the concessionaire, a payment is made at the end of the concession to compensate for the difference between the projected traffic levels used in the above calculations and the real traffic level observed. This compensation is calculated as

(box continues on following page)

Contract design should be as specific as possible with respect to such ongoing adjustments as inflation, so that these risks are handled routinely. Project risks and uncertainty in the economic and financial environment, however, will inevitably create situations that

Box 6.5 continued

$$R_{N+S} = \sum_{i=k+1}^{N+S} \frac{Y_t - T_t(\hat{Y}_t - \hat{T}_t)}{(1+r)^{1-k}} * (1+\rho)^{N+S-k}$$

where the "^" symbol indicates the ex post real value observed of the variable. No compensation exists, however, for operational and maintenance costs that differ from the original estimates. Otherwise, the concessionaire would have an incentive to inflate these costs in order to receive extra compensation at the end of the concession period. These costs are usually small in comparison to investments, however. Differences between the expert's estimate used to calculate the compensation and the real ex post costs are unlikely to have a significant effect on the concession's profitability.

Source: Gómez-Lobo and Hinojosa (2000).

require contract renegotiation. The concession contract should specify the conditions that would allow renegotiation of the contract terms, the types of events that could trigger renegotiation, and the frequency with which reviews can occur. The contract also should specify what remedies are available to the regulator for restructuring, for example, whether concession length might be extended or an investment program might be modified. Too often one initiates contract renegotiation for a specific issue and then expands it to other issues. This approach is prone to corruption and creates incentives for sponsors to seek contract revisions on a regular basis.

- *Dispute resolution.* The agreement should explain the procedures for settling disputes in a fair and timely manner, including provisions for arbitration or mediation. Foreign concessionaires may request that such disputes be resolved in a neutral jurisdiction. Peru recognized this in its recent draft contracts, and it now always includes a clause explaining how disputes will be settled and when international arbitration will be used. In a nutshell, an expert (picked randomly if the parties cannot agree to one) will resolve technical conflicts locally, and an international arbitration commission will resolve nontechnical conflicts over a certain amount. Below that amount, they are resolved locally. All local decisions are made within specific time limits.

In Chile the main dispute settlement mechanism is the Conciliatory Commission. This commission has three members, one nominated by the concessionaire, one by the authorities, and one by mutual accord. Commission members must be nominated at the beginning of the concession before any controversies have arisen. The commission is established when one of the parties raises a demand. In the case of the state, contracts stipulate an explicit and limited set of circumstances in which it can raise a demand to the commission. The concessionaire has more flexibility in this respect. The commission's initial task is to conciliate the diverging positions. If an agreement is not reached, the concessionaire, and only the concessionaire, has the choice of either taking the matter to the judicial system or requesting the establishment of an Arbitration Commission. The same members of the Conciliatory Commission form this last commission, and its decision is binding and not subject to appeal in the courts.[11]

Toll Road Auctions and Award Criteria

As in most infrastructure sectors, competition in the road sector is essentially *for* the market. Because the toll franchise has a degree of exclusivity, the auction is a crucial element to help ensure that services are being provided efficiently. Given the complexity of road infrastructure projects and the diversity of objectives that road agencies tend to have for their projects, coming up with an ideal bidding rule is often difficult for governments. Table 6.4 shows the diverse approaches that have been used. Many countries have adopted a two-stage process in which they evaluate technical proposals separately from and prior to financial proposals. They then select the winning bidder from those that pass the technical evaluation.

While technical validation helps reduce the risk of project failure, it may also have important drawbacks. It often involves considerable discretion and judgment by the evaluation committee, which reduces the overall transparency of the process. Experience also has shown that changing market conditions after the contract award may require operators to make significant changes to the project. These changes reduce the meaningfulness of the initial technical evaluations to the extent that they rely on the base forecasts.

To remedy this, many governments are issuing a preliminary set of technical standards to be achieved, which is subject to discussion and modification with prospective bidders. This has been Chile's experience (Gomez-Lobo and Hinojosa 2000). Interaction often takes place with the regulator, which is desirable because the regulator will eventually be responsible for monitoring compliance. After this consultation, the bidding package is finalized so that the parties bid on the same technical specifications and requirements and the winner is picked from the financial proposal. This is wonderful from a regulator's viewpoint, because if enough

11. For more details on Chile see Gomez-Lobo and Hinojosa (2000).

Table 6.4. Award Criteria in Selected Latin American Toll Road Concessions

Country	*Award criteria*	*Concession duration*
Argentina—road corridors	Highest lease fee paid to government	Fixed by government but extended after renegotiation
Argentina—urban access	Lowest toll	Fixed by government but extended after renegotiation
Brazil—Federal	Lowest toll	Fixed by government
Brazil—Sao Paulo	Highest lease fee paid to government	Fixed by government
Brazil—Parana	Largest network length	Fixed (but likely to be extended as a result of politically imposed cut in toll)
Chile—1st generation	Multiple criteria	Fixed by government
Chile—2nd generation	Least net present value	Unknown
Colombia—1st generation	Multiple criteria	Fixed by government
Colombia—2nd generation	Least cost to government	Fixed by government
Mexico	Shortest term	Fixed by bid
Peru	Shortest term	Fixed by bid
Peru	Least subsidy	Fixed by government
Uruguay	Shortest term	Fixed by bid

Sources: Irigoyen (1999); various World Bank internal reports.

potential bidders participate in the discussion and the various bidders do not collude, the process converges toward what could be referred to as consensus engineering cost. The regulator now has some idea of what best practice investment, maintenance, and operation costs should be for a specific road.

This is not the end of it. Many different options still exist for structuring financial proposals for road concessions. Some of the more common include (a) the lowest toll level, (b) the shortest duration of the concession, (c) the highest payment to the government for existing infrastructure, and (d) the lowest subsidy that the government requires.

Less common options include the lowest income guarantee that the government requests and the amount of new investment or its speed, as well as some innovative ideas discussed later. As regulators learn about past mistakes, the way that toll roads are being auctioned evolves.

THE INITIAL EXPERIENCES WITH COMPETITION FOR THE TOLL ROADS MARKET. The earliest road concessions (such as the first generation of Argentine and Chilean toll roads) were trying to be everything to everyone and were awarded following complex, weighted, multiple criteria picked from the list just described. This was a source of opaque and often subjective, if not corrupt, decisionmaking. Next, when governments started to see that simpler is better and decided to focus on a single criterion, bidding tended to be based either on the minimum toll (as in the second generation of Argentine toll roads) or, if the toll was specified, the shortest duration for the franchise (as in the initial Mexican toll road program). Both these approaches presented significant incentive problems. Bidding on the basis of the minimum toll may result in poor price signals in congested corridors. If one sets tolls exclusively to cover investment, maintenance, and operating costs, then high tolls result when low traffic volumes are expected and low tolls result in high traffic and congested conditions.

Similarly, bidding based on the shortest concession period also has problems, especially if tolls are not specified. In Mexico, where projects were tendered based on the shortest concession duration that firms offered for a given traffic flow, shorter concession durations necessitated the setting of higher tolls in order to finance the projects. The resulting high tolls produced important traffic diversions and many complaints to regulators from users with a limited ability to pay. Mexico's requiring alternative freeways for each concession did not help. The ultimate outcome was a financial situation so catastrophic that it required a subsequent government bailout for many roads.

Bidding on the basis of investment commitments has been used to develop road networks, but this also has had problems that often result in operator demands for renegotiation. By locking in future investment levels, the concessionaire is prevented from adjusting investment to meet changing market conditions. Second, it may encourage overoptimism and excessive investment (see box 6.6).

THE NEW, IMPROVED COMPETITION FOR THE TOLL ROADS MARKET. In the wake of the bailouts, new schemes have been developed to improve incentives and reduce the risks of road concessions. In Peru bidding has taken place in terms of the minimum amount of required government investment in each concession. This serves to buy down the size of the project and reduce the financial risk exposure of the concessionaire. In the United Kingdom, the design-build-finance-operate scheme establishes the government payment of shadow tolls based on traffic volumes. This provides a long-term mechanism for government support that phases out as traffic volume grows.

Chile has developed perhaps the most innovative road concession programs—although it is facing its fair share of problems with many of the contracts being renegotiated in 1999–2000. As a reaction to the low bidding

Box 6.6. *Why Were Consortia Initially So Optimistic About Road Projects?*

Regulators also need to understand the motivation behind the optimism, because in many cases the outcome of excessive optimism is renegotiation. Indeed, many of the earlier road concessions have experienced problems. Concessionaires have been either overly optimistic or overly aggressive in bidding, leading to a host of restructuring and renegotiations. Firms pursue this strategy for several reasons, namely:

- A "first mover" advantage to grab exists when several projects are going to be concessioned. By winning the first bid, firms signal their low cost or aggressive behavior to other bidders, with the goal of discouraging future competition.
- Because construction firms are often the key consortium partner, construction contracts rather than the subsequent operation of the concession are the dominant interest, and bidding below cost secures the construction contracts, with disregard to the long-term financial viability of the concession, which will be the problem of the other consortium members or the creditors.
- Firms may bid low just to win the franchise with the sincere intention of renegotiating the contract as soon as possible. Few governments have refused to renegotiate. Indeed, if the concession runs into financial problems in the future, associated political problems occur as well as costs and delays in retendering the project. Therefore, bidding low and renegotiating afterward may be a viable strategy for a potential concessionaire (a phenomenon called "lowballing").
- Finally, one cannot rule out optimization mistakes on the part of bidders, possibly related to poor assessment of demand uncertainty ("winner's curse"), or the complexity of tendering mechanisms.

Source: Based on Gomez-Lobo and Hinojosa (2000).

problem, Chile tendered its Route 5 Temuco–Rio Bueno concession on the basis of a minimum toll, within a band set by the government. The floor of the band is set sufficiently high to guarantee a minimum revenue stream to the concessionaire. In addition, the bidding documents fix the contract duration. Setting this minimum toll level and the contract duration effectively puts a floor on the concession company's expected earnings. Therefore, the risk of future financial distress for the concession firm (which would force the government to renegotiate the contract) is minimized—although not eliminated, as seen in recent developments. If two or more firms bid the minimum value, the winner is the one that offers the highest transfer directly to the government.[12]

12. Because this transfer does not affect the concession firm's income or capital structure, sponsors can bid as much as they like without jeopardizing the concession's financial stability. If investors make a mistake and bid too much, the consequent loss will show up in the financial returns of the sponsor, not the concession company. For more details see Gomez-Lobo and Hinojosa (2000).

In effect, this bidding mechanism significantly reduces the chance of renegotiation, but it does not lower the competitive pressure of the process.[13] If the concession firm is in good shape and no risks of disruption to its activities exist because of financial distress, governments should be better equipped to resist renegotiation pressures. This transfer mechanism from sponsors has served to generate close to US$150 million in the four concessions where it has been used. The proceeds are deposited in an infrastructure fund that is then used to cross-subsidize other projects or pay for minimum income guarantees.

THE NEWEST FORMS OF COMPETITION FOR THE MARKET. Chile also has pioneered, at least academically, another bidding approach that holds some promise for dealing with the fixed-term nature of traditional franchising contracts and that Colombia and Mexico are considering for both roads and airport runways.[14] The bidding variable, instead of toll levels or another conventional variable, is the present value of revenue throughout the life of the concession that firms are willing to accept to undertake the project. The firm that bids the lowest present value of revenue wins. The duration of the concession is then flexible and depends on the effective traffic levels encountered. Once the concessionaire has received (in present value terms) the amount that it bid, the concession ends and the infrastructure reverts to public ownership. If real traffic levels are lower than expected, the concession duration is extended automatically, while if traffic is higher than expected, the opposite occurs. Therefore, income uncertainty due to traffic variations is largely eliminated for the concessionaire.

In addition, the LPVR auction reduces potential conflicts related to the early termination of a concession. In a 10- to 30-year contract, excessive traffic growth or other events may occur that require added investments. Canceling the original contract and retendering the concession with the extended projects would be optimal, rather than negotiating the additional investments with the existing concessionaire. This seldom happens, because it

13. Engel, Fisher, and Galetovic (1997a,b) in several of their articles point out that the lowest present value of revenue (LPVR) auction may also reduce the occurrence of lowballing. Their argument rests on the assumption that the winning firm's LPVR bid offers the government a credible threat to terminate the concession quickly and compensate the firm if it tries to renegotiate. This is discussed later, in relation to the LPVR auction mechanism.

14. The United Kingdom was the first country to apply a variable-length concession with an LPVR flavor in the Severn, Trent, and Dartford bridge concessions.

would require a difficult estimation of compensation for the forgone future income stream owed for an early contract termination. The LPVR auction reduces this problem substantially by giving the concessionaire the difference between what it originally bid and what it has already earned. From a regulator's viewpoint, another important characteristic of the LPVR mechanism is that tolls can be adjusted without having to negotiate new terms with the concessionaire. If tolls are deemed too high or low, the authorities could change them without affecting the concessionaire's expected income stream and without engaging in a potentially protracted negotiation process. As stressed in Engel, Fischer, and Galetovic (1997a,b), this flexibility may be important in urban road concessions in which determining the optimal tariff ex ante is difficult, especially during congestion periods.

The LPVR mechanism also has its drawbacks. It may lower the incentive of concessionaires to make demand-enhancing investments such as quality improvements. The increase in demand from these expenditures results in an earlier termination of the contract, with little benefit to the concessionaire.[15] Perhaps a more important difficulty is that the LPVR auction does not resolve possible cash flow problems that a concessionaire may face when traffic levels drop.

Another limitation occurs in cases in which operation and maintenance costs are relatively high compared with construction costs. A low-traffic situation then puts the concessionaire in trouble, because the extension of the contract generates increasingly high maintenance costs that eventually may make the project unsustainable. Hence, although the risk of demand is reduced under LPVR, it is not completely eliminated. Bidders still have to estimate the future level of traffic to compute their required revenue. A possible way to refine the LPVR mechanism is to require bidders to provide separate offers for construction and average annual operating costs. (For more details on this refinement, see De Rus and Nombela 1999.)

Price Regulation

One of the main reasons why toll projects fail is that privatization teams have a hard time assessing demand prospects. In turn, one of the main reasons why demand prospects are hard to assess is that traffic levels often

15. Early termination of the contract would save the concessionaire the additional maintenance and operation costs that would have been incurred during the original period, but these are usually small.

depend on what economists call the elasticity of demand with respect to price—in other words, how sensitive demand is to changes in prices. In practice, this matters a lot, particularly in developing countries where the ability to pay is often limited and regulators are sometimes expected to make recommendations based on the social impact of pricing decisions. This explains why so many differences in toll design and toll levels can be found across countries. In principle, they have to reflect costs, but the specific costs to be covered can vary (construction/rehabilitation, operations, maintenance, environmental, safety, and congestion costs). In general, the toll calculation reflects the first three types of costs, and the last two are beginning to be incorporated. Environmental costs have tended to be included only to the extent that they entitle the operator to specific recoverable expenditures.

Table 6.5 shows that in general, countries tend to fix the toll levels needed to recover investment, operation, and maintenance costs. It shows that the price cap is now a common form of regulation in the sector, just as in many of the others. The last column suggests that in many cases, governments end up restructuring the toll levels at some point (jointly with subsidies to the toll operators or the extension of a contract term). These contractual changes are such that price caps are transformed into rate of return regulation, because the main purpose of the adjustment is to shift part of the risk imposed on the operator through a price cap back to the users (through longer contracts) or to the government (through subsidies).

Table 6.5. *Toll Design and Levels in Selected Latin American Toll Road Concessions*

Country	*Toll design*	*Per km car rates (in U.S. cents)*	*Restructuring needed*
Argentina—road corridors	Fixed	1.56	Yes
Argentina—urban access	Capped	3.5	Yes
Brazil—Federal	Capped	2.3–5	Yes
Chile—1st generation	Capped	2–3	No
Colombia—1st generation	Fixed	3–4	No
Mexico—public toll roads	Fixed	2–11	Yes
Mexico—private toll roads	Fixed	13–50	Yes
Uruguay	Fixed	3.5	Yes
Venezuela	Fixed	1	No

Sources: Irigoyen (1999); and various World Bank internal reports.

One of the main concerns that road operators have to address and regulators have to understand is the uncertainty about introducing direct pricing in the sector. In defining price regulation, the following challenges must be tackled:

- *Question 1: How much should the operator recover through the toll system?* More specifically, what is the level investment the operator should be allowed to recover, given current and forecasted traffic levels? Unfortunately, this investment is a moving target, because roads tend to alternate between excess capacity at off-peak times and congestion and capacity shortfalls at peak times. This problem also arises in a longer-term sense. Indeed, because building road capacity takes time, what appears to be excess capacity today may meet demand in five years' time. Also, it makes sense, to minimize costs over time, for an operator and a government to take some bets (for example, a four-lane bridge may only cost 50 to 60 percent more than a two-lane bridge). This also is, of course, often a political challenge that a regulator has to justify, because opposition to tolls is sometimes based on the excess capacity observed at the beginning. One solution for minimizing the perception of overcharging is to allow the operator to look for alternative sources of financing from subconcessions such as gas stations, restaurants, playgrounds, or advertising, but these seldom yield much more than 5 to 10 percent of the revenue needed.
- *Question 2: Should tolls be fixed or should they vary greatly during the lifetime of the investment?*[16] This is a complex regulatory question with multiple dimensions and viewpoints.
  - The economist's answer will be that when a facility first opens, the optimal price will be close to zero, or at least very low, because it only needs to cover operation and maintenance. The road, which was sized for future traffic growth, will be uncongested in the early years and hence have a negligible marginal cost. Later, as traffic builds up, congestion and the optimal road price will grow as well. But when traffic reaches the maximum and new road capacity is added, the optimal price will again fall sharply.

16. Economists refer to this problem as the difference between short-run and long-run marginal costs.

- The typical politician's answer will be to keep the toll stable. The vast majority of fixed tolls identified throughout Latin America, as seen in table 6.5, reflect the domination of this position around the world. The consequence of fixed tolls is that road operators tend to overcharge in the early years and undercharge later.[17]
- The economic problem is, in practice, under control, because many contracts now have toll escalation clauses (generally subject to regulatory approval, but in some cases automatic) that are related to a local consumer or construction price index, increasingly calculated in dollars, to offset the potential effects of a devaluation.
- In addition, one increasingly recognizes that the option to price congestion is a good one, and this eases the possibility of future toll increases. In many countries, peak and off-peak tolls are already different. In the longer run, a larger share of the day will end up being considered peak time and hence ease the recovery of revenue needed to cover higher maintenance costs resulting from higher traffic. Once more, the ideal arrangement for a regulator is to ensure compliance with formula-driven adjustments.
- The answer to this question depends largely on the amortization rules the road operator is allowed. If, for whatever fiscal reason, the operator can follow a fast-track amortization for investment in a road, the toll will be high at the beginning and lower once the road is amortized fiscally, because the only expenses left to recover are operation and maintenance. The regulator in this sector is responsible for monitoring and possibly defining these amortization rules, as in most of the other sectors. Without clear rules, using loose amortization rules is one of the instruments for operators to argue for toll increases to strategically distribute costs over time.

- *Question 3: Should the regulator require tolls to be differentiated across users?* One can consider several dimensions to differentiation.
  - *The different road damages that different vehicles impose.* This arises because automobiles and trucks impose different requirements on roads (see box 6.7 for a technical explanation), which is why

17. Debt service requirements concentrated in the middle years of a facility's life make this pricing problem worse. Such financing burdens are even greater for developing countries with limited access to long-term capital markets.

Box 6.7. *A Brief Lesson in Engineering for the Price Regulator*

Costs, and hence toll differentiation, should be driven by the demands that different vehicles place on the shared road. One can divide highway costs into two types: the basic capacity to carry traffic and pavement durability and smoothness. Civil engineers measure the demands that different vehicle types place on capacity relative to that of a standard passenger car (known as PCEs, or passenger car equivalent units). The number of PCEs of capacity needed by a heavy truck varies according to terrain and other factors. For example, on a level road, a truck may only represent the equivalent of 1.2 cars, while on a moderate-grade road, the lower horsepower-to-weight ratio of trucks might make them the equivalent of 4 passenger cars. The number of lanes, lane width, grades, curves, and other factors determine the traffic-carrying capacity of a road. Pavement durability is determined by the type of pavement, its thickness, and the stresses to which it has been subject since construction.

Road damage is a function not of the size of the vehicle but of the weight being borne on the axles of a vehicle. Road engineers measure the demands that different vehicle types place on roads in terms of the damage caused by the passage of a reference axle weighing 18,000 pounds, approximately the weight on axles of many heavy trucks. This is known as an equivalent standard axle load, or ESAL. Pavement damage increases at the third or fourth power of axle weight, so that the 1,000-pound axle loading on a typical car produces only about 1/10,000 the pavement damage of a typical heavy truck axle. This nonlinear damage impact is offset to some degree by the fact that the number of ESALs that a road can withstand before it needs to be resurfaced or rebuilt is a power function of pavement thickness. For example, a pavement that is 11 inches thick is about twice as durable as one that is 9 inches thick, yet it costs only a fraction more to build.

These can be major issues from a private operator's viewpoint, because car volumes dominate carrying capacity metrics and truck characteristics and volumes are key to pavement durability aspects. In essence, cars tend to be responsible for the number of lanes, while trucks are responsible for how thick each lane should be. In economic terms, road charges should have two components: one for pavement damage, based on ESALs; and one for congestion, based on PCEs. In practice, though, road pricing tends to use total weight rather than axle loadings for trucks, and attempts at congestion pricing based on PCEs are only beginning (such as in California, Singapore, and the United Kingdom).

most concession contracts allow at least some degree of differentiation between cars and trucks and buses. In practice, unit tolls will vary according to the number of axles on the vehicle as an approximation for the wear and tear each vehicle imposes on the road's pavement. In general, trucks and buses are charged two to four times the level of automobile tolls, with the precise amounts varying depending on size, weight, traffic mix, and development objectives.

- *The political viability of differentiating tariffs across regions of a same country.* Many governments impose the same tolls/kilometer across a country, because some politicians find it difficult explaining that interregional toll differences can be justified by differences in construction and maintenance costs. As explained earlier, this means that explicit subsidies may sometimes be required or that cross-subsidies need to be tolerated for some operators, but this also means that regulators must have access to sufficiently detailed cost data to ensure that there is no abuse and that users are not overcharged.

- *Social pricing.* In many poor countries around the world, the main inter-urban roads are likely to be important infrastructures for rural users who need to take their products to urban centers. This represents serious social concerns, as well as strong interest groups with political clout. This may be why governments commonly impose a special treatment of some user groups and have their use of the roads financed through some type of shadow toll or subsidy/voucher. Peru is considering this solution, for instance, to address farmers' protests against the tolling of some highways. Similarly, for urban access roads, allowing lower tolls for public transportation users makes sense, because often these are likely to include the poor. In addition, using buses reduces congestion and pollution.

- *Question 4: How high can a toll really be?* A limit exists, of course. A user's willingness to pay tolls is a function of income, the value assigned to time savings, reductions in vehicle operating costs, and the cost and quality of competing alternatives. On average, toll rates have ranged from US$0.01 to US$0.10/kilometer/car. Some congestion-related tolls in Europe and the United States run between US$0.15 and US$0.20/kilometer/car, while some bridge and tunnel tolls may range up to US$0.50 /kilometer. Special situations also can be found in which toll levels are far above these averages (the result of high costs and legal requirements for inflation adjustments). In Mexico, tolls have risen to more than US$0.60 /kilometer in a couple of cases. This experience clearly shows that a ceiling exists to the willingness to pay, because in these cases, traffic volumes have tended to be low relative to capacity. These concessions frequently have encountered severe financial problems, and congestion on alternative roads has not really been alleviated. Eventually, regulators have been forced to accept renegotiation.

- *Question 5: How much freedom should the operator be allowed to differentiate its toll structure?* As long as the operator stays within the allowed rate of return or overall price cap, and as long as no competition exists (such as predatory pricing aimed at capturing business from a competing mode on a specific road), the regulator has no reason to interfere with a tariff structure design aimed at making the most of user willingness to pay or at expanding the regular customer base. An operator can design its tariff structure in many ways to achieve these goals and maximize profits:

  - *Variation by time of day:* This is commonly allowed.
  - *Congestion pricing:* This is becoming increasingly popular.
  - *High-speed lanes:* The idea of allowing the price of one lane to change with the degree of congestion to service users in a rush allows the operator to make the most of differences in the various users' value of time.
  - *Discounts for loyal customers:* Tolling technology is now allowing the recognition of a frequent user basis among commuters; and to promote the growth of these clients, some companies are proposing special discounts to well-targeted groups, including local residents or car pool commuters. In Argentina, for instance, the users of an electronic toll get a discount on some segments. Their prices vary from US$0.80 to US$1.10, compared with the normal toll of US$1.40 to US$1.50.

Quality Regulation

For concession performance, an economic regulator must be concerned with three main quality issues: the technical quality of the road, compliance with contractual obligations, and safety and environmental issues.

The Technical Quality of Roads

One needs to consider road quality issues at the outset of concession design and technical specification. Technical matters such as pavement materials, thickness, and construction techniques must be specified from the beginning, because these aspects will help determine the facility's performance and future maintenance and investment needs. An inventory of the initial state of assets is a minimum requirement for effective economic regulation.

Asset quality indicators include pavement roughness and deterioration, condition of lighting, markings, signaling, quality of fire and rescue equipment, condition of maintenance, and weather-related equipment (such as snowplows). The monitoring, inspection, and certification of the initial construction and investment is essential and should include all related investments such as signage, pavement markings, toll collection facilities, fire and rescue services, and access points. It also may extend to ancillary and support facilities such as service stations and restaurant plazas.

Once the toll road is in operation, quality aspects shift to ensuring that the assets are maintained, that performance standards are achieved, and that additional investments are made when performance triggers are reached. Performance standards, which should be established in the original concession agreement, should include asset quality, operating conditions, safety indicators, and emergency readiness.

The regulatory authority should be prepared to audit records and inspect equipment on a regular basis. It also needs to ensure that the concessionaire has reserved sufficient funds for maintenance and repair of the assets. This is particularly problematic in the later years of the contract, when incentives to maintain equipment and facilities are lower. Also, if subcontractors provide any of these services, the regulatory authority should be able to monitor the contract terms and the financial capability of all parties to the contract.

In practice, what regulators generally do in an increasing number of countries is to match performance against the established parameters and quality standards set in the World Bank Highway Management Program. This is effective enough to identify performance outliers for most technical variables.

Operating Quality of Road Services

While quality service requires asset maintenance, the regulator should not forget that the goal is to provide transport services worth paying for. The operating performance of the system is central to public support and to determine at what point additional investment may be required. The concession contract should establish performance standards that cover the following:

- Lane availability and shutdowns
- Traffic volumes and average speeds, both peak and off-peak
- Toll queue performance: waiting times and availability
- Capacity, speed, and visibility during inclement weather
- Access conditions and bottlenecks

- Activity levels at service plazas
- Response times and service aspects of emergency vehicles.

The concessionaire should be required to provide data on these performance aspects on a regular basis (monthly or quarterly), subject to review and audit. If actual performance is below the standard, the contract should specify the nature and type of sanctions to be imposed or the nature and timing of investments to be undertaken in response. This can be tricky in practice. For example, not meeting a performance standard concerning the length and time in toll queues could be the result of traffic growth (requiring new investment) or poor maintenance of collection equipment (requiring improved performance by the concessionaire). This issue is particularly important when new investment requires revisions to the concession contract.

Safety Aspects

Safety regulation takes a number of different forms. First, the facility itself must be designed to handle the anticipated traffic volume and mix under a variety of operating conditions. These dimensions include such technical factors as capacity, speed, grades, roughness, signaling, lighting, and emergency services.

Safety is not only a function of the physical characteristics of the road, however, but also the quality and operation of the vehicles using the road. In particular, speeding, unsafe driving practices, and poor vehicle inspection practices can lead to accidents. Most road concessions, however, rely on existing police and motor vehicle registration/inspection services provided by the government, usually on a reimbursement basis. Here again, performance standards can help evaluate whether safety problems are the result of the facility or are from traffic enforcement shortcomings. For example, if average speeds are above the statutory limit, this may indicate reduced or ineffective enforcement.

Another aspect of safety involves vehicle standards, especially truck size and weight requirements. Because revenues from trucking activity are critical to toll road viability, how trucking regulation is handled is important. In some cases, concessionaires operate truck inspection and weighing stations; in other cases, public authorities handle them on a reimbursement basis. Problems have arisen when stricter enforcement of weight regulations (overloading) has led truckers to avoid toll roads. If enforcement is relaxed, however, this leads to a much faster rate of pavement deterioration and, in many cases, higher accident frequency and severity.

Overall, safety aspects of toll roads should be built in to design and operating standards. In practice, though, the nature of traffic and vehicle enforcement in the country will shape accident rates and safety performance. Because of this, countries should consider toll road initiatives as providing an opportunity to improve public safety throughout the road network.

Environmental Aspects

Environmental issues first emerge in contract design during siting and planning decisions and must take into account geography, construction techniques, and the facility's operating practices. Initially, mitigation measures could include adapting designs with respect to alignments, materials used, and standards for construction. During construction, the concession should specify particular investments required to improve environmental aspects, including noise barriers, retention ponds, and other remedial measures, as well as relocation and resettlement issues, if they arise. During both construction and operation, the regulatory authority should ensure compliance with environmental laws, including such aspects as use of salt and chemicals, runoff, and recycling of pavement materials. In practice, the environmental agency rather than the road authority often controls this aspect, although the two institutions have obvious interactions.

Working with User Feedback

In addition to monitoring assets and performance, road concessions should have a mechanism for public participation and feedback. This can be handled through a regular system of surveys as well as through the designation of a user group that can provide information on the qualitative aspects of the concession. Because toll roads tend to be highly visible, and in many cases controversial, designing mechanisms for public input is important for evaluating performance, to extend public knowledge of the project, and to build public support.

This has to be handled with some care, however. In Brazil, for instance, each concession is required to survey customer satisfaction every six months or so. Overall, user satisfaction with the toll roads has been quite positive, although it is deteriorating. The problem is that these results are biased because users have already demonstrated their belief in the value of the toll road by using it and continuing to use it increasingly, despite what appears to be a worsening of satisfaction. Indeed, users who were extremely satisfied with the immediate improvements in quality in the first year of road operation quickly forgot about the initial conditions of the road and started to focus on their unhappiness at having to pay for a (bad) road that used to be free. People are noticing improvements, but they are also being managed by customer service improvements, which have nothing to do with road services. Special events for children, presents for drivers, and similar campaigns just before the surveys can be effective in managing the emotions of the toll road users at the right time.

Performance Indicators and Information Requirements

We now have the new economics of private road concessions. We have learned about the extent to which road concessions are vulnerable to macroeconomic conditions, exchange rate shocks, and income growth. Demand has proven sensitive to toll levels, income, gross domestic product, and trade activity. These sensitivities, along with incentives to "buy in the deal and then figure out how to make it work" on the part of sponsors and creditors, have meant that many private toll roads have required public financial support. The challenge is to design new structures that take into account the reality of public-private linkages and a more activist role for the public sector in monitoring and regulating concessions.

Regulators will not be able to work on such a structure unless they have enough information. Once more, the contract has a key role to play in this context. Road concession contracts should contain an annex that specifies specific reporting requirements (including clear definitions) for the concessionaire, the frequency of the reporting requirements, and their format to facilitate comparisons across projects. This information is required not only to monitor contract compliance, but also to identify when additional investments are needed and to help resolve disputes.

Too often, public authorities have placed great emphasis on technical specifications and sponsor prequalification in contract design, but pay less attention to making sure they have good, timely information about the performance of the concession and the sponsor. However, even in situations in which technical and operating information is consistently supplied, governments have been faced with problems emerging from heavily leveraged projects or from weak sponsor balance sheets. In principle, nonrecourse project financing of toll roads should place primary emphasis on the economics of the project itself. The need for more equity capital in toll road projects means that profits from construction activities are not enough, so that both project cash flows and the sponsor's financial condition must be stronger than in the past. The 1990s saw a large number of construction company bankruptcies, however, so sound projects may be at risk because of weak sponsors. This could occur through a lack of investments being made due to a shortage of funds, to losses on other projects reducing the sponsor's equity capital, or to financial risks from exchange rate or refinancing.

Thus, regulatory authorities need a range of technical, operational, and financial information not only about the project, but about the project participants themselves. This information is not intended to be used to micromanage the concession, but rather to serve as an early warning system to reduce the likelihood and costs of restructuring and bailout. Such information should include the data shown in table 6.6.

The project operational indicators are intended to monitor the physical aspects of the project, from pavement and equipment conditions to performance in terms of facility availability, safety, and technical efficiency. The revenue indicators are intended to monitor the performance of the contract, and they are especially important when the government is providing revenue or traffic guarantees. Reporting of revenues across time periods and by user groups is needed to understand the structure of demand and to monitor sponsor efforts to raise revenues through discounting and so forth. Revenues from ancillary services are needed to understand the basis for rate of return or price cap regulation and to understand the interaction between direct toll and ancillary revenues. Cost data are needed to make sure that services are being provided at the lowest cost and to monitor costs to be included in regulatory calculations.

Because many road concessions are designed to bring new investment, information is needed about ongoing investment activity compared to contract requirement and budget plans. This investment information should be compared with traffic volumes to validate prior forecasts and to determine whether these programs should be delayed or accelerated.

Project financial indicators are intended to monitor the liquidity, solvency, and profitability of the concession. These indicators are similar to those contained in covenants that creditors impose and are needed for rate of return or price cap regulation.

Regulators have tended to underappreciate the importance of detailed information about the quality and value of assets. First, the quality of assets is central to the performance of the road. Second, the long-lived nature of road infrastructure means that deferring maintenance is relatively easy in the short run to boost returns, allowing road and equipment to deteriorate and accelerate major overhaul requirements. Third, the treatment of

Table 6.6. Reporting Requirements for Road Concessions

Concession operating performance	Lane availability Average speed by time of day Toll station availability Toll station queueing time by time of day Accident and safety indicators Availability of emergency equipment Engineering quality indicators (roughness, signage, lighting)
Revenue indicators	Traffic volume by vehicle class Traffic volume by time of day (peak/off-peak) Revenue collected by vehicle class Revenue collected by time of day (peak/off-peak) Revenue generated by ancillary services Revenue from enforcement levies Revenue and volumes from different discount programs (commuter, high frequency)
Cost indicators	Operating expenses by activity: Toll collection Road maintenance Road operations Emergency services Cost of special services for particular users (for example, truck weigh stations)
Investment indicators	Investment spending vs. budget (including variance analysis) Physical investment (for example, lane-km resurfaced)
Project financial indicators	Profit as percent revenues Working capital Debt service coverage Debt service projections Debt-equity ratio Debt-assets ratio Return on assets Return on equity
Assets by class (gross and net of both tax and regulatory amortization)	Road infrastructure Equipment Ancillary services Maintenance and renewal program
Sponsor financial information	Income statement, balance sheet and cash flow statements (audited) Working capital Debt service schedule and currency structures

Source: Authors.

asset depreciation and amortization is important in determining the base from which rates of return are computed. In general, tax policies typically allow the write-off of road infrastructure on an accelerated basis or with a shorter tax life than economic life. This disparity means that after the facility is depreciated for tax purposes, incentive is reduced to maintain the asset. Moreover, if regulatory accounting for the concession uses a longer amortization period than tax accounting, the higher regulatory net asset values at any point in time will lead to higher tolls to provide a specified rate of return on assets. The sponsor thus receives higher returns in the early years from tax depreciation, then higher returns from the regulatory accounting that includes asset valuations already written off for tax purposes. Thus governments need either to harmonize regulatory and tax treatment of assets or to make sure that regulatory rate of return calculations take into account tax benefits from accelerated depreciation.

Finally, ongoing financial reporting is needed by the sponsors themselves, beyond the specific project. The government should require sponsors to provide audited financial statements to ensure that the prequalification status is maintained throughout the life of the concession. Debt servicing schedules and working capital positions should supplement standard financial statements. These data will help ensure that the specific project is not put at risk by financial troubles of the parent, a twist on the traditional concerns of nonrecourse financing.

Conclusions: Recent Innovations and Emerging Issues

We know a lot more about the challenges involved in getting highways and urban access roads tolled than we did at the beginning of the 1990s. All players know much more about it: the sponsors are more prepared to face risks that they understand better, the construction companies are probably even more anxious to get involved, and the users have also learned about how to fight for their rights more effectively. Regulators also know much more, but their knowledge is more an appreciation of how little they have known about the business of monitoring toll road packages prepared by consultants or privatization teams. Regulators have also learned that they will often be firefighters and that the only way to be effective in that role is to better prepare for the job while they can. Too few toll road regulators have been successful in improving their preparation, but with a little help from this chapter, the hope is that they will be asking more and better questions of the teams that are preparing toll road packages, and that they will be able to argue more effectively with concessionaires who are trying to renegotiate commitments to the government.

References

ADB (Asian Development Bank). 1999. *Developing Best Practices for Promoting Private Sector Investment in Infrastructure: Roads.* Draft for book.

Alvarado, R. 1998. "La organización de la estiba en Centroamérica y países del Pacto Andino." *I Congreso Internacional de Estiba.* Las Palmas de Gran Canaria, Spain: Puertos del Estado.

Anuario del Transporte. 1997. Basauri (Vizcaya), Spain: Transeditores, S.A.

Armstrong, M., S. Cowan, and J. Vickers. 1994. *Regulatory Reform, Economic Analysis and British Experience.* Cambridge, Massachusetts: MIT Press.

Ashford, N., and P. H. Wright. 1992. *Airport Engineering.* New York: John Wiley & Sons, Inc.

Baumol, W. J. 1977. "On the Proper Cost Tests for Natural Monopoly in a Multiproduct Industry." *American Economic Review* 65(5): 810-22.

_____. 1995. "Modified Regulation of Telecommunications and the Public Interest Standard." In M. Bishop, J. Kay, and C. Meyer, eds., *The Regulatory Challenge.* Oxford: Oxford University Press.

Baumol, W. J., J. C. Panzar, and R. D. Willig. 1982. *Contestable Markets and the Theory of Industry Structure.* Jokanovich, New York: Harcourt Brace.

BIE (Bureau of Industry Economics). 1995. *International Benchmarking: An Overview.* Report no. 95/20. Canberra: Australian Government Publishing Service.

_____. 1995a. *Rail Freight 1995: International Benchmarking.* Report no. 95/22. Canberra: Australian Government Publishing Service.

_____. 1995b. *Issues in Infrastructure Pricing*. Research Report no. 69. Canberra: Australian Government Publishing Service.

Borenstein, S., and M. B. Zimmerman. 1988. "Market Incentives for Safe Commercial Airline Operation." *American Economic Review* 78(December): 913-35.

Borts, G. H. 1960. "The Estimation of Rail Cost Functions." *Econometrica* 28(1): 108-131.

Boyer, K. D. 1997. *Principles of Transportation Economics*. Reading, Massachusetts: Addison-Wesley.

Bradshaw, B. 1996. "The Privatization of Railways in Britain." *Japan Railway and Transport Review* 8(September): 15-21.

Braeutigam, R. 1980. "An Analysis of Fully Distributed Cost Pricing in Regulated Industries." *The Bell Journal of Economics* 11(1): 182-96.

_____. 1989. "Optimal Policies for Natural Monopolies." In *Handbook of Industrial Organisation*, vol. II. Amsterdam, New York: North Holland.

Brennan, T. 1995. "Mexican Government Buys Out Union Contracts to Speed Port Privatization." *Traffic World* 242 (June).

Button, K. J. 1993. *Transport Economics*. New York: Edward Elgar Publishing.

_____. 1994. "Privatization and Deregulation: Its Implications for Negative Transport Externalities." *The Annals of Regional Science* 28(1): 125-38.

Burns, P., and A. Estache. 1998. "Information, Accounting, and the Regulation of Concessioned Infrastructure Monopolies." Policy Research Working Paper no. 2034. World Bank, World Bank Institute, Washington, D.C.

Carbajo, J. C., and A. Estache. 1996. "Railway Concessions—Heading Down the Right Track in Argentina." *Viewpoint 88*. Finance and Private Sector Development Department. Washington, D.C.: World Bank.

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

Cass, S. 1996. *Port Privatisation: Process, Players, and Progress*. London: IIR Publications.

Chalk, A. 1983. "The Economics of Airline Safety." Ph.D. diss., Washington University, St. Louis, Missouri.

Christensen, P. 1998. "Internalization of Externalities." *Deliverable 7: Pricing European Transport Systems (PETS)*. Project No. ST-96-SC.172. Research project for the Commission of the European Communities, Brussels.

Crampes, C., and A. Estache. 1998. "Regulatory Trade-Offs in the Design of Concession Contracts." *Utilities Policy* 7(1): 1-13.

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

De Rus, G., and G. Nombela. 1999. "Variable Term Concessions for Highways: Are the Most Efficient Concessionaires Always Being Selected?" World Bank Institute, Washington, D.C.

De Rus, G., C. Román, and L. Trujillo. 1994. *Actividad Económica y Estructura de Costes del Puerto de La Luz y de Las Palmas.* Madrid: Ed. Cívitas.

De Rus, G., M. Gonzalez, C. Román, B. Tovar, and L. Trujillo. 1995. "Competitividad de los Puertos Españoles." Working paper. Tribunal de Defensa de la Competencia, Madrid.

Dodgson, J. 1994. "Access Pricing in the Railway System." *Utilities Policy* 4(3): 205-213.

Doganis, R. 1991. *Flying Off Course: The Economics of International Airlines.* London: Harper Collins.

_____. 1992. *The Airport Business.* London: Routledge.

Doganis, R., and A. Graham. 1995. "The Economic Performance of European Airports." Research Report no. 3, University of Cranfield, Cranfield, U.K.

DSC (Drewry Shipping Consultants). 1998. *World Container Terminals: Global Growth and Private Profit.* London.

ECMT (European Conference of Ministers of Transport). 1996. *Evolution des Transports* (1970-1994). ECMT/OECD, Paris.

Ecole Nationale des Ponts et Chaussees. 1998. International Symposium: Road Financing, Construction, Maintenance and Operation, November 4-6, Paris.

Engel, E., R. Fischer, and A. Galetovic, 1996. "Licitación de carreteras en Chile." *Estudios Público* 61(Summer): 5-37.

_____. 1997a. "Highway Franchising: Pitfalls and Opportunities." *American Economic Review, Papers and Proceedings* 87(2): 68-72.

_____. 1997b. "Privatizing Roads: A New Method for Auctioning Highways." *Viewpoint* 112. Washington, D.C.: World Bank. Reprinted in S. Smith, ed., *The Private Sector in Infrastructure.* Washington, D.C.: World Bank.

_____. 1998. "Least-Present-Value of Revenue Auctions and Highway Franchising." Working Paper no. 6689. National Bureau of Economic Research, Cambridge, Massachusetts.

ESPO (European Sea Ports Organisation). 1996. "Report of an Enquiry into the Current Situation in the Major Community Sea Ports." Fact Finding Report, revision 1996. European Commission, Brussels.

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

Estache, A., and J. Carbajo. 1996. "Designing Toll Road Concessions—Lessons from Argentina." *Public Policy for the Private Sector* 99. Washington, D.C.: World Bank.

Estache, A., J. Carbajo, and G. De Rus. 1999. "Argentina's Transport Privatization and Re-regulation: Ups and Downs of a Daring Decade-Long Experience." Policy Research Working Paper no. 2249. World Bank, Washington, D.C.

Estache A., J. Campos, N. De Castro, and J. Strong. 1999. "Peru: Getting Ready for Successful Public-Private Partnership in Transport." World Bank, Washington, D.C.

Estache A., J. Campos, G. Nombela, J. Strong, and L. Trujillo. 1999. "Mexico: Facing New Challenges for Public-Private Partnership in Transport." World Bank, Washington, D.C.

Estache A., I. Alexander, J. Campos, G. Nombela, J. Strong, and L. Trujillo. 1999. "Brazil: Public-Private Partnership in Transport Infrastructure—Progress and Challenges." World Bank, Washington, D.C.

European Commission. 1993. *European Economy: Annual Economic Report for 1993.* Brussels.

European Commission. 1997. *Green Paper on Sea Ports and Maritime Infrastructure.* Official Publications of the EU Communities, COM (97) 678 final. Brussels.

European Parliament. 1993. *European Sea Port Policy.* Directorate General for Research, Transport Series E-1, 7-1993. Luxembourg.

Fayard, A. 1993. "Toll Financing, Risk Financing, How to Fit the Needs Without Dogmas: The French Experience." Presented at the East-West European Road Conference, September 22-24, Warsaw.

Ferrer, J. 1997. "Assessment of U.K. Port Privatisation." Ph.D. diss., Department of Maritime Studies and International Transport, University of Wales, Cardiff.

FHD (French Highway Directorate). 1999. "Analysis of Highway Concessions in Europe." Report for the Western European Road Directors. Paris.

Fishbein, G., and S. Babbar. 1996. "Private Financing of Toll Roads." RMC Discussion Paper No. 117. Washington, D.C.: World Bank.

Fleming, K., and A. Ghobrial. 1993. "A Normative Framework for Assessing Airport Security." *Transportation Quarterly* 47(3): 413-19.

Forsyth, P. 1984. "Airlines and Airports: Privatization, Competition and Regulation." *Fiscal Studies* 5(February): 61-75.

_____. 1997. "Price Regulation of Airports in Australia." Department of Economics, Monash University, Clayton, Australia. Paper presented at the Air Transport Research Group Conference, June 25-27, Vancouver.

Friedlander, A., E. R. Berndt, J. S. Chiang, M. Showalter, and C. A. Vellturro. 1993. "Rail Costs and Capital Adjustments in a Quasi-Regulated Environment." *Journal of Transport Economics and Policy* 27(2): 131-52.

Fujimori, Y. 1997. "Ten Years of JNR's Privatization (Technological Development after JNR's Reformation." *Japanese Railway Engineering* 37(1-2): 3-5.

Galenson, A., and L. S. Thompson. 1993. "Forms of Private Sector Participation in Railways." *TWU Papers*, no. 9. Washington, D.C.: World Bank.

Gannon, C., and Z. Shalizi. 1995. "The Use of Sectoral and Project Performance Indicators in Bank-Financed Transport Operations." *TWU Papers*, no. 21. Washington, D.C.: World Bank.

Gathon, H. J., and P. Pestieau. 1995. "Decomposing Efficiency into its Managerial and its Regulatory Components: The Case of European Railways." *European Journal of Operational Research* 80(3): 500-507.

Gómez-Ibañez, J. A. 1997. "Sri Lanka Transport: The Bus Industry." Case CRI-97-1377.0. Kennedy School of Government, Harvard University, Cambridge, Massachusetts.

_____. 1999. "Pricing." In J. A. Gomez-Ibanez, W. B. Tye, and C. Winston, eds., *Essays in Transportation Economics and Policy*. Washington, D.C.: The Brookings Institution.

Gómez-Ibañez, J. A., and J. R. Meyer. 1993. *Going Private: The International Experience with Transport Privatization.* Washington, D.C.: The Brookings Institution.

Gómez-Lobo, A., and S. Hinojosa. 2000. "Broad Roads in a Thin Country: Infrastructure Concessions in Chile." Policy Research Working Paper no. 2279. World Bank, Washington, D.C.

Goss, L. 1990. "Economic Policies and Seaports: 3. Are Port Authorities Necessary?" *Maritime Policy and Management* 17(4): 207-219.

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

Guislain, P., and M. Kerf. 1995. "Concessions—The Way to Privatize Infrastructure Sector Monopolies." *Public Policy for the Private Sector* 59. Washington, D.C.: World Bank.

Gwilliam, K. 1998. *EDI Regulation Course.* Washington, D.C.: World Bank.

Häfner, P. 1996. "The Effects of Railroad Reform in Germany." *Japan Railway and Transport Review* 8(September): 27-30.

Hall, K. G. 1998. "Port Privatization." *The Journal of Commerce* 416(May 1): 1B.

Halperin, R., and P. Malone. 1999. "Overview of World Experience in Private Financing in the Road Sector." Infrastructure Sector Unit Working Paper no. 2. World Bank, Washington, D.C.

Hanke, S. H., and R. K. Davis. 1973. "Potential for Marginal Cost Pricing in Water Resources Management." *Water Resources Research* 9(4): 808-825.

HAPFP (Highways Agency and Private Finance Panel). 1997. *DBFO-Value in Roads: A Case Study on the First Eight DBFO Road Contracts and Their Development.* London: Central Office of Information.

Hayuth, Y., and D. Hilling. 1992. "Technological Change and Seaport Development." In B. S. Hoyle and D. A. Pinder, eds., *European Port Cities in Transition.* London: Belhaven Press.

Heggie, I. G., and P. Vickers. 1998. *Commercial Management and Financing of Roads.* Technical Paper no. 409. Washington, D.C.: World Bank.

HMSO (Her Majesty's Stationery Office). 1986. *The Airport Act.* London.

_____. 1998. *The Private Finance Initiative: The First Four Design, Build, Finance and Operate Road Contracts*. London.

Hoffman, J. 1998. "What Are the Obstacles, Advantages, and Disadvantages of Port Privatizations? It All Depends!" Economic Commission for Latin America and the Caribbean, Santiago, Chile.

ICAO (International Civil Aviation Organization). 1991. *Airport Economics Manual*. Doc 9562.

Irwin, T., M. Klein, G. E. Perry, and M. Thobani. 1997. *Dealing with Public Risk in Private Infrastructure*. Washington, D.C.: World Bank, Latin American and Caribbean Studies.

ITS (Institute for Transport Studies). 1994. "European Railway Comparisons." Final Report, Working Paper no. 418. University of Leeds, U.K.

Jansson, J. O. 1984. *Transport System Optimization and Pricing*. New York: John Wiley and Sons.

Juhel, M. 1994. "Port Development and Management in Central and South America." 28th Congress, Permanent International Association of Navigational Congresses, Sesión Iberoamericana (Permanent Commission for Development and Cooperation), Seville, Spain.

_____. 1997. "Government Regulation of Port Activities: What Balance between Public and Private Sectors?" *International Course on Privatization and Regulation of Transport Services*. Washington, D.C.: World Bank.

Kapur, A. 1995. *Airport Infrastructure: The Emerging Role of the Private Sector*. Technical Paper no. 313. Washington, D.C.: World Bank.

Kent, P. E., and A. Hochstein. 1998. "Port Reform and Privatization in Conditions of Limited Competition: The Experience in Colombia, Costa Rica and Nicaragua." *Journal of Maritime Policy Management* 25(4): 313-33.

Kerf, M., R. D. Gray, T. Irwin, C. Lévesque, and R. R. Taylor. 1997. *Concessions for Infrastructure: A Guide to Their Design and Award*. Technical Paper no. 399. Finance, Private Sector, and Infrastructure Network. Washington, D.C.: World Bank.

Kessides, I. N., and R. D. Willig. 1995. "Restructuring Regulation of the Rail Industry for the Public Interest." Policy Research Working Paper no. 1506. World Bank, Washington, D.C.

Klein, B., and K. B. Leffler. 1981. "The Role of Market Forces in Assuring Contractual Performance." *Journal of Political Economy* 89(4): 615-41.

Kopicki, R., and L. S. Thompson. 1995. *Best Methods of Railway Restructuring and Privatization.* Cofinancing and Financial Advisory Services Discussion Paper Series no. 111. Washington, D.C.: World Bank.

Kuranami, C. 1999. "Review of Recent Toll Road Development in Selected Countries." Paper presented at the Asian Toll Road Development in an Era of Financial Crisis Seminar, March 3, Tokyo.

Laffont, J. J., and J. Tirole. 1993. *A Theory of Incentives in Procurement and Regulation.* Cambridge, Massachusetts: MIT Press.

Layard, R., and S. Glaister, eds. 1994. *Introduction to Cost-Benefit Analysis.* Cambridge, U.K.: Cambridge University Press.

Lemaitre, A. 1997. "The Measurement of Airport Performance." Unpublished. Institute of Transport Studies, University of Sydney.

Levesque, T. J. 1994. "Modelling the Effect of Airport Noise on Residential Housing Markets." *Journal of Transport Economics and Policy* 28(May): 199-210.

Liston, C. 1993. "Price-Cap Versus Rate-of-Return Regulation." *Journal of Regulatory Economics* 5(1): 25-48.

Loudon, R. M. 1984. "Region of Durham Experiences in Pricing and Water Conservation." *Canadian Water Resources Journal* 9(4): 19-29.

Lundberg, A. 1996. "Restructuring of the Swedish State Railways." *Japan Railway and Transport Review* 8(September): 22-26.

Matoba, J. I. 1999. "Toll Road System in Japan: Expressway Network Development Through Public Corporations." Paper presented at the Asian Toll Road Development in an Era of Financial Crisis Seminar, March 3, Tokyo.

Mercer Management Consulting. 1996. "Private Financing of Toll Roads: A Review of Experience." World Bank, Washington, D.C.

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

MMC (Monopolies and Mergers Commission). 1996. *A Report on the Economic Regulation of the London Airports Companies: Heathrow Airport Ltd, Gatwick Airport Ltd and Stansted Airport Ltd.* No. MMC4. London: The Stationery Office, Ltd.

_____. 1997. *A Report on the Economic Regulation of Manchester Airport plc.* No. MMC5. London: The Stationery Office Ltd.

Morrison, S. A. 1982. "The Structure of Landing Fees at Uncongested Airports." *Journal of Transport Economics and Policy* 16(May): 151-59.

Mumayiz, S., and N. Ashford. 1986. "Methodology for Planning and Operations Management of Airport Terminal Facilities." *Transportation Research Record* 1094: 24-35.

Nash, C. 1985. "European Rail Comparisons: What Can We Learn?" In K. J. Button and D. E. Pittfield, eds., *International Railway Economics*. Aldershot, U.K.: Gower.

Nash, C. A. 1996. "Separating Rail Infrastructure and Operations: The British Experience." European Conference of Ministers of Transport Round Table No. 103, June 13-14, Paris.

_____. 1997. "Rail Privatisation: How Is It Going?" Institute for Transport Studies Working Paper no. 497. University of Leeds, U.K.

Nelson, J. P. 1980. "Airports and Property Values." *Journal of Transport Economics and Policy* 14(1): 37-52.

Nombela, G., L. Trujillo, and A. Matas. 1998. "The Economic Aspect of Charging for Trans-European (TEN) Seaport Infrastructure in the European Union." Final Report, European Commission, Brussels.

ORR (Office of the Rail Regulator). 1994. *Railtrack's Track Access Charges for Franchised Passenger Services: Developing the Structure of Charges: A Policy Statement.* London: Her Majesty's Stationery Office.

_____. 1995. *Framework for the Approval of Railtrack's Track Access Charges for Freight Services.* London: Her Majesty's Stationery Office.

_____. 1997. *Licensing of Rail Operators.* London: Her Majesty's Stationery Office.

OSC (Ocean Shipping Consultants Ltd.). 1996. "The World Container Port Market to 2010: Competition, Investment, Capacity Utilisation and Prospects." *Ports and Harbors* 41(4): 18-19.

Oum, T. H., and C. Yu. 1994. "Economic Efficiency of Railways and Implications for Public Policy." *Journal of Transport Economics and Policy* 28(2): 121-38.

Papon, F. 1998. "Comparaison des Principes de Financement des Routes dans Differents Pays." In *Le Financement de La Route*. Paris: Presses de l'École Nationale des Ponts et Chaussées.

PDE (Puertos del Estado). 1998. "Recopilación y análisis de tarifas y otros ingresos en puertos extranjeros." Study of A. T. Kearney for the Spanish Public Agency of State Ports, Madrid.

Peters, H. J. 1995. "Private Sector Involvement in East and Southeast Asian Ports: An Overview of Contractual Arrangements." *TWU Papers*, no. PS-10. Washington, D.C.: World Bank.

Piron, V. 1999. "Urbain-Interurbain: La Problematique est devenue globale." *Transports* 393(January-February): 25-42.

Poole, R. W. 1990. "Airport Privatization—What the Record Shows." PS124(August). The Reason Foundation, Los Angeles.

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

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

PWF (Public Works Financing International). 1998. *International Major Projects Survey*. Westfield, New Jersey.

Raciatti, J. L. 1998. "La organización de la estiba en los países de Mercosur: Chile y Argentina." *I Congreso Internacional de Estiba*. Las Palmas de Gran Canaria, Spain: Puertos del Estado.

Rose, N. L. 1990. "Profitability and Product Quality: Economic Determinants of Airline Safety Performance." *Journal of Political Economy* 98(5): 944-64.

Ruster, J. 1997. "A Retrospective on the Mexican Toll Road Program (1989-1994), The Private Sector in Infrastructure: Strategy, Regulation, and Risk." *Public Policy for the Private Sector* 125. Washington, D.C.: World Bank.

Saunders, R. J., J. J. Warford, and P. C. Mann. 1977. "Alternative Concepts of Marginal Cost for Public Utility Pricing: Problems of Application in the Water Supply Sector." Paper no. 1259. World Bank, Washington, D.C.

Shapiro, C. 1982. "Consumer Information, Product Quality and Seller Reputation." *Bell Journal of Economics* 13(1): 20-35.

Shaw N. L., K. M. Gwilliam, and L. S. Thompson. 1996. "Concessions in Transport." *TWU Papers*, no. 27. Washington, D.C.: World Bank.

Slack, B. 1985. "Containerisation, Interport Competition and Port Selection." *Journal of Maritime Policy Management* 12(4): 293-303.

Small, K., and C. Winston. 1999. "The Demand for Transportation: Models and Applications." In J. A. Gomez-Ibanez, W. B. Tye, and C. Winston, eds., *Essays in Transportation Economics and Policy.* Washington, D.C.: The Brookings Institution.

Smith, W. 1996. "Utility Regulators: Creating Agencies in Reforming and Developing Countries." Paper presented at the International Forum for Utility Regulation, June 21, London. Reprinted in World Bank. 1997. *Public Policy for the Private Sector.* Washington, D.C.

Stigler, G. J. 1971. "The Theory of Economic Regulation." *The Bell Journal of Economics* 2(1): 3-21.

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

Stiglitz, J. 1998. "The Private Use of Public Interests: Incentives and Institutions." *Journal of Economic Perspectives* 12(2): 3-22.

Stopford, M. 1997. *Maritime Economics.* London: Ed. Routledge.

Sutter, M. 1997. "Privatized Ports Gain Amid Growing Pains." *The Journal of Commerce* 411(March 4): 6A.

Suykens, F. 1996. "Influence of Port Tariffs on Maritime Transport." Paper presented at VI International Congress of Maritime Traffic and Port Handling, Vigo, Spain.

Swift, J. 1997a. "Regulating the Railway in the Public Interest." Paper presented at Office of the Rail Regulator Railway Study Association Sessional Meeting, June 11, London School of Economics, London.

_____. 1997b. "Regulation of the Railways in Great Britain: 1993-1997 to 2000+." Paper presented at Office of the Rail Regulator, Address to the European Union Delegation of the French National Assembly and the Senate, March 13, Paris.

Talley, W. K. 1988. *Transport Carrier Costing.* New York: Gordon and Breach Science Publishers.

Tenenbaum, B. 1996. "Regulation: What the Prime Minister Needs to Know." *The Electricity Journal* 9(2): 28-36.

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

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

Thompson, L. S., and J. M. Fraser. 1996. "Financial Success Hinges on Productivity." Database, Rail Business Report 1995. World Bank, Washington, D.C.

Tortello, M. 1998. "La organización de la estiba en los países de Mercosur y Chile: El modelo portuario de Chile." *I Congreso Internacional de Estiba.* Las Palmas de Gran Canaria, Spain: Puertos del Estado.

Transport Canada. 1996. *Commercialization of the Air Navigation System, The Canadian Experience.* Ottawa.

Trujillo, L., E. Quimet, and A. Estache. 2000. "Forecasting the Demand for Privatized Transport: What Economic Regulators Should Know and Why." World Bank, Washington, D.C.

Valletti, T., and A. Estache. 1999. "The Theory of Access Pricing: An Overview for Regulators." Policy Research Working Paper no. 2097. World Bank, Washington, D.C.

Viegas, J., and C. Fernández. 1997. "Review of Current Situation." *Deliverable: Pricing European Transport Systems (PETS).* Project No. ST-96-SC.172. Research project for the Commission of the European Communities, Brussels.

Walters, A. A. 1975. *Noise and Prices.* New York: Oxford University Press.

_____. 1978. "Airports: An Economic Survey." *Journal of Transport Economics and Policy* 12(2): 125-60.

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

Webster, M., and P. Cohen. 1997. "Ports: Continued Buoyancy for Private Ports." *Infrastructure Yearbook 1997.* London: Ed. J. Wade.

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

_____. 1997a. *Private Participation in Infrastructure.* Washington, D.C.

_____. 1997b. *Toolkits for Private Participation in Water and Sanitation.* Washington, D.C.

World Bank Toll Road Website. Accessed August 17, 1999. www.worldbank.org/html/fpd/transport/roads/toll_rds.htm

Wright, C. L., and D. J. Freire Coloma. 1997. "Toll-Road Partnerships: What Works, What Doesn't, and Why?" *Transportation Quarterly* 51(4): 85-101.

Index

(Page numbers in italics indicate material in figures, tables, or boxes.)

Continued from inside front cover

Other WBI Development Studies
(In order of publication)

Monitoring and Evaluating Social Programs in Developing Countries: A Handbook for Policymakers, Managers, and Researchers
Joseph Valadez and Michael Bamberger

Agroindustrial Investment and Operations
James G. Brown with Deloitte & Touche

Labor Markets in an Era of Adjustment
Edited by Susan Horton, Ravi Kanbur, and Dipak Mazumdar
Vol. 1—Issues Papers; Vol. 2—Case Studies

Does Privatization Deliver? Highlights from a World Bank Conference
Edited by Ahmed Galal and Mary Shirley

Financial Regulation: Changing the Rules of the Game
Edited by Dimitri Vittas

Privatization and Control of State-Owned Enterprises
Edited by Ravi Ramamurti and Raymond Vernon

Finance at the Frontier: Debt Capacity and the Role of Credit in the Private Economy
J. D. Von Pischke

To order World Bank Institute (WBI) publications

by phone or fax:
Phone: 1-800-645-7247 or (703) 661-1580
Fax: (703) 661-1501

by mail:
The World Bank, P.O. Box 960, Herndon, VA 20172-0960, U.S.A.

by e-mail:
books@worldbank.org

through the World Wide Web at:
www.worldbank.org/html/extpb/Publications.html

The World Bank

1818 H Street, N.W.
Washington, D.C. 20433, U.S.A.
Telephone: (202) 477-1234
Facsimile: (202) 477-6391
Telex: MCI 64145 WORLDBANK
MCI 248423 WORLDBANK
Internet: www.worldbank.org
E-mail: books@worldbank.org

"I was impressed by the scope and depth of the manuscript. The sector-specific chapters can be easily understood because they provide many examples focusing on particular industries."

Prof. Jose A. Gómez-Ibáñez
Harvard University

"The quality of the book is excellent. It combines very effectively a theoretical analysis with a presentation of facts."

Prof. Emile Quinet
Ecole Nationale des Ponts et Chaussées (Paris)

9 780821 347218 14721

ISBN 0-8213-4721-7