

Regulatory trade-offs in the design of concession contracts

Claude Crampes *

Antonio Estache **

Preliminary version

Comments Welcome

August 1997

* IDEI and Gremaq, Université des Sciences Sociales de Toulouse
Place Anatole France, 31042 Toulouse Cedex, France
tel: 33 561 12 86 02 fax: 561 12 86 37
email: crampes@cict.fr

** World Bank and ECARE, 1818 H street, Washington, D.C. 20433, U.S.A.
tel: 1-202-458 14 42 fax: 1-202-334 83 50 email: aestache@worlbank.org

We benefited from comments from O. Chisari, G. de Rus, R. Green and M. Rodriguez-Pardina. Any mistake is ours. Also, the findings, interpretations, and conclusions are our own and should not be attributed to the World Bank, its Executive Board or any of its member countries.

Abstract

The paper discusses how conflicts between the multiple objectives of policymakers (efficiency, equity, fiscal, speed of reform, signaling,...) can influence the optimal design of concessions contract for network services in infrastructure. The discussion covers the relevance of information asymmetries for the optimal regulatory regime and processes, for the optimal design of awarding processes and for the optimal duration of the contracts. It also addresses the design of pricing in relation with the universal service obligation. It concludes with an illustration of the preceding economic principles with some stylized facts from the liberalization experiment in Argentina.

1. Introduction

Network activities typically involve the collection of a good or service from a large number of producers and/or its distribution to a large number of users. Moreover, both producers and users are often scattered on a wide geographical area. In such industries, economies of scale and of scope as well as benefits from pooling and coordinating are invoked to justify a close financial integration. And since it is well recognized that huge private firms could behave in a way that satisfies neither the economist (because of lack of efficiency) nor the non-economist (because of lack of equity), in many countries network operators are completely integrated publicly-owned firms. However, during the last two decades, this point of view has been challenged by the idea that the lack of efficiency resulting from a lack of competition overcompensates the gain from economic integration.

Competition starts with the unbundling of the various stages of the delivery process. The type and degree of competition to introduce used to be justified by the classical trade-off between internal and external efficiencies¹. When economies of scale are not too strong with respect to the size of the market, the risk of both types of inefficiency is weak and actual competition can be expected to be sufficient to promote overall efficiency. This is the case for example in electricity generation where competition can work through a free-entry rule both on the supply and demand sides². Concerning now the transport and distribution activities, if the good or service needs a continuous material infrastructure to be conveyed from one point to another (pipes, wires, roads...), "standard" competition is very unlikely to be the first-best solution since it would imply a multiplication of the infrastructure in the same neighborhoods. Duplication costs of the infrastructure would offset the social benefits from the presence of a competitor. But on the other hand, a monopoly is naturally induced to an external misallocation of resources. How can the government solve this dilemma?

Even when a monopoly seems to be the best solution to deliver a service, competitive mechanisms can still work. A popular approach among policy advisors in developing countries and increasingly in OECD countries is the competitive bidding of the concession of services by the public authority to a private monopolistic provider. The authority can keep the property rights on the production assets and exclusively concede their operation to a private firm. The winner of the competitive awarding mechanism obtains the right to behave as a profit-maximizer in so far as it respects some quality and environmental qualifications and under some redistribution obligations (for instance, universal service and no price-discrimination). As a residual claimant for the current

¹ Internal efficiency refers to the relative choice of inputs by firms, while external efficiency refers to their sales and pricing policies.

² In Spain, power supply is allegedly opened to any producer ... with an authorization from the Ministry of Energy. On the demand side in Argentina, access to the wholesale market is restricted to large customers. But by the legal definition, the customers authorized to intervene are less and less large; they first had to buy a capacity larger than 10 MW and, after successive decreases, the limit was fixed at 100 kW on February 1995.

profit, the private monopoly has a strong incentive to minimize operating costs. And if the bidding process has resulted in an efficient access fee to get the concession, the funds collected by the government can be reallocated in a way that satisfies its distributive objective as well.

But in all liberalization processes, efficiency is not the only element to be taken into account. Social and fiscal feasibility can be first rank concerns that will impede the pure economic solution. The paper discusses how concessions contracts for the management of network services can be allocated and monitored and how trade-offs are created when governments are trying to meet too many objectives through a concessioning strategy. Specific examples are given, mostly drawn from the electricity and water industries. In the next section, we present the diversity of the regulatory objectives for a government facing a private monopoly. In section 3, we recall the information asymmetries of the government and we explain why a concession contract can be a second best way to approximate the optimal mix of public concerns. Section 4 is dedicated to the problem of pricing in relation with the universal service obligation. Section 5 focuses on the risks of capture of the regulatory agency. In section 6, we study the awarding process of the concession contracts. Finally, in section 7, we illustrate the preceding economic principles with some stylized facts from the liberalization experiment for the delivery of water and electricity in Argentina.

2. Regulatory mechanisms and the diversity of objectives

The point of view of most economists on the regulation of economic activities has significantly evolved over the last three decades.³ The circumstances under which a monopoly can be more efficient than competition are well understood.⁴ The idea of a monolithic and myopic upright ministry perfectly regulating a monolithic greedy and non reactive private firm has been dropped as well. Both the principal (the government) and the agent (the firm) are now viewed as complex entities with strategic behaviors and using their private information as scarce valuable resources.

On the other hand, efficiency concern which is the core of the economic models is often very far from the political plans despite some apparent convergence. Consider the case of the monopoly. Both the economist and the politician are against private monopolies but for different reasons. The economist does not like monopolists because they do not provide enough good or service to the market, which results in deadweight loss of surplus. In this negative judgment, the distribution of surplus between the consumers and the firm does not matter at all. As for the government, a monopoly is bad because it chooses too high prices and so it appropriates unacceptable profits from consumers who *do* buy from it.

When the economist presses the firm to increase its output and the government urges the firm to decrease its price, they seem to agree implicitly. Actually, this is only a circumstantial convergence of their purposes and it is easy to find solutions that satisfy one of them without

³ See Laffont-Tirole (1993)

⁴ The monopolists is more efficient for instance to benefit from large scale operations or to internalize environmental damages.

satisfying the other. For instance, politicians have little trouble with a price cut without any change in the output level. In this case, the economist disagrees since it creates a rationing without any improvement in the social surplus.⁵ On the contrary, the economist can solve the inefficiency monopoly's curse by allowing it to fix perfect discriminating prices. Doing so, the monopolist will produce the socially optimal level of output since it can extract the whole surplus from consumers. But the politician will be furious against this solution on the grounds that (i) everyone should pay the same price and (ii) it is too much in favor of the firm.

Consequently, while they do agree on the necessity to regulate monopolies, the politician and the economist differ on how to do it. The economist will favor any solution that promotes efficiency ... provided it is feasible. Because of this last qualification, the economist can accept public monopolies or regulated private monopolies despite its preference for competitive mechanisms. Symmetrically, the politician is essentially interested in the households' welfare in so far as it conditions his\her re-election. Consequently, he\she prefers solutions that are inexpensive for the public budget (taxpayers' concern) and that protect the social equilibrium (no social exclusion but not too much price discrimination). Then he\she can opt for some market solutions even if he prefers to keep some control on the operation and development of the firm.

These compromises may explain why the concession system is increasingly of interest to researchers on the provision of public services. In a concession system, a firm is allowed to use a part of the public assets for a given duration in network industries. The key element of the effective regulation of this firm is the design of the contract, starting with the description of the rights and obligations of the concessionaire. The private operators get to use public assets such as water, air or land but at a lower political cost than pure privatization since large shares of the population view these as assets that can only belong to the national patrimony and are happy to see the state retain their ownership. The concession arrangement is then a useful compromise between purely public and purely private organizations of a network industry.

The specific design of the concession system, its regulatory regime, the tariff regime or even the way the concession is awarded reveal the priorities of the government, most obviously the ranking of fiscal vs. static and dynamic efficiency concerns. Most reasons that motivate the intervention of the government in a given industry are simultaneously present in a specific regulatory process but their ranking is difficult because they represent the concerns of very different interest groups.⁶ From an economic normative viewpoint, all concerns should be introduced in the decision process as subordinated to efficiency concerns, or at least as exogenous constraints that define the feasible set wherein the most efficient solution will be chosen. In practice however, efficiency is often considered by public authorities as a second order concern as compared to income distribution and budgetary concerns. The following sections show how these diverse and sometimes antagonistic constraints and objectives interact and can and should

⁵ In France, during the 60s and 70s, it was not uncommon to wait for 2-3 years before obtaining the installation of a telephone line.

⁶ Among the reasons, we can enumerate to raise fiscal revenue, to promote efficient static and dynamic choices, to modify the distribution of income, to increase the quality and safety of products, to protect the environment,...

influence the design of the payment scheme of the regulated service provider.⁷

3. Rewards, incentives and risk-sharing

In starting the design of the regulatory regime for a concessionaire, the government has to address three essential questions. It has to decide i) how much the regulated firm has to pay the government for the use rights received through the concession, ii) how much the regulated firm should be paid to cover the cost of providing the service (including some profit), and iii) how to collect the money for these payments. The interactions between the answers given to these three questions provide the basic elements of the incentive structure faced by the concessionaires for the duration of the concession but they also determine the extent to which the participation of the firm will be a binding constraint for the government under various levels and types of risks. For instance, will the concessioned service package be attractive to the private sector at all, or will it be most attractive to the best in the field? These three issues are addressed next.

Incentive constraints

The cost coverage and the net profit payment to the service provider can flow from two sources: from the market and from any kind of public subsidy. Clearly, how the reward is computed and the money collected will induce a strategic behavior by the firm because managers have an informational advantage over the regulator. Since management efforts are costly, a high-tech firm has an incentive to pretend it is less efficient. Imitating inefficient firms allows the incumbent to avoid the disutility of management efforts. So, a subsidy or a pricing rule exclusively based on the observation of ex post costs cannot be efficient since it would result in the potentially productive firm performing a bad management effort. More generally, any rewarding system that gives the firm a fixed net remuneration, such as cost-plus payment, tends to generate poor incentives to improve upon an existing situation. This is particularly undesirable in activities where technological improvements are significant, as in telecommunications.

The conclusion is clear from an efficiency viewpoint: the payment should be based on a rule which allows for at least some degree of variation of the net payment with the observed ex post cost. Among the large family of variable payments, the so-called two-part payment is appealing by its performance and its simplicity. It combines a fixed price or transfer (to be sure that the firm will perform the project, as we will see later) and a variable price or transfer (to induce it to manage the project efficiently).

⁷ The standard optimization approach in public economics is to maximize the social welfare subject to a large set of constraints: legal, fiscal, technological, informational etc. In practice, the regulatory mechanisms contain all these concerns without a clear difference between the statutes of objective or of constraint. When several are introduced as simultaneous objectives, it can be the result of a rivalry between several ministries. The modern approach of the regulation by multi principals analyses this problem. See Martimort (1992).

But it is not sufficient for the government to propose a unique two-part payment with a variable part based on the observed cost: doing so, the government would deprive itself of the use of information available within the firm. For the regulatory mechanism to be efficient, the regulator has to use all the information present in the industry: not only the cost that will be observed ex post but also the information privately owned by the managers. How can this be achieved? The idea is to reward the firm on the basis of the information it chooses to disclose concerning its technology and/or demand. The reward mechanism gives the right incentives if the firm earns higher profits when reporting the true value of its technology rather than lying.

Specifically, an efficient firm should report it is efficient and an inefficient one that it is inefficient. In short, the optimal non linear payment must be different if one can observe ex post low or high cost and if the firm reports low or high cost. A simple way to do that is to give the firm a payment decreasing with its reported cost every time the observed cost is equal to the announced one and to force it to pay very high penalties when the two costs are different.

Public subsidization, risk sharing and participation

But this kind of incentive-based mechanism is not necessarily the most effective in meeting the fiscal concerns of the government. Indeed, if the efficient mechanism implies that the government has to keep on allocating significant resources to the service it concessioned, it may not get the political support required for its implementation when the main priority of the reforming government is to shrink the annual deficit due to subsidies given to network activities. Indeed, when the revelation of information has a fiscal cost because of some type of explicit subsidy, the government has to raise funds to finance activities it is trying to leave to the private sector. And, as long as the government has to raise taxes, efficiency can be threatened since the shadow price of public funds is fairly high.

These general principles remain valid when cost can be affected by random events, i.e. when commercial, regulatory, political or any other type of risks are involved.⁸ Here, we only consider how the ex ante contract should be designed when the production cost is affected by exogenous shocks. The first way to deal with these shocks is to try to anticipate them and to write a complete contract, which means a contract depicting the rights and duties of both parties in each conceivable state of nature. This is why for concessions, the conditions of contracts are usually long and precise. But a contract can never be complete, first because there are too many imaginable contingencies (state of the technology, price of inputs, regulatory environment, etc.)

⁸ Another extension is when the production can vary in size and/or quality. Then, the contract should be made contingent on them... provided they are verifiable. Because of the information asymmetry on technical characteristics and effort, in the optimal regulation mechanism there is an under-provision of effort and of quality/quantity; so the same will happen in any non linear contract mechanism. Usually, measuring the quantity variable is technically easy, but it can be very costly to collect exact measures at the individual level when consumption is scattered, like in the water, gas or electricity industries. Concerning quality, when the contract is devoted to the procurement of a good to the government, the civil servants can check whether or not it conforms to admitted standards. It is more difficult to get information on quality when the firm produces a marketed good.

and many contingencies are unforeseeable, second because the cost of monitoring them or the cost of getting their enforcement by a court is too high.

An incomplete contract just specifies what should be done when unforeseen events do occur. Parties can decide to renegotiate from the very beginning or to negotiate only on the new contingencies, to give equal bargaining powers to both parties or to concede leadership to one party, to call for a third party arbitration, etc. For the economist, the best solution is the one that does not discourage any effort to enhance efficiency. For instance, if the project needs some preliminary specific investment as for most infrastructure projects, the firm should have some leadership in the renegotiating process to prevent an under-investment that would occur if the regulator were the one who can decide when and how to renegotiate. For the politician, this efficiency concern is also present in so far as it can govern the continuity of the service provision, which is a condition for the satisfaction of voters.

As a consequence of the incompleteness of contracts, the government and the firm are facing a risky situation and this also affects the optimal choice of the regulatory mechanism. If both are risk neutral, it does not matter who the risk bearer is. When the manager is risk adverse while the regulator remains risk neutral, besides the cost observability problem, it will often be optimal to transfer some or all the risks to the regulator or to the consumer giving the firm a reward independent of the ex post observed variables.

When risk is a serious concern and the policy-makers need to rely on the private sector, the regulatory mechanism is not only limited by incentive constraints but also by participation constraints. Indeed, if there are some chances that ex ante and ex post costs strongly differ depending on exogenous events uncontrolled by the concessionaire, the firm may simply not agree to participate when penalties for non-delivery are high or when it has to take on all of the risks by itself. This is particularly true in developing or transforming economies where relative prices tend to change dramatically during the adjustment process and where non-commercial risks can appear overbearing for potential entrants.

If the regulator decided to design a payment mechanism totally depending on the firm's effort and if even with the best of efforts, the activities can only result in financial losses, no firm would be interested, despite the social desirability of production. Often, to ensure participation, concerns for risk will have to dominate concerns for efficiency. In other words, a cost-plus or rate-of-return regulatory regime can be the only way to ensure that private operators are interested in participating. But doing so, we give the firm full insurance on its remuneration and destroy the incentive effect of a variable reward.

Menu of contracts and hierarchy of concerns

The second best strategy for the government is to offer a menu of so-called incentive contracts, each composed of a fixed payment, a function of the announced cost, and a linear sharing of overruns, that is, of the difference between ex post observed costs and announced costs. This menu approximates the optimal regulatory mechanism, which trades off the truthful

elicitation of information about productivity (which would lead to a cost plus contract) and the ex post inducement of an appropriate level of effort (which would lead to a fixed-price contract). Within this menu, the firm will choose the contract that maximizes its profit and this self-selecting process is optimal since choosing one contract is like revealing internal information. A firm that expects a large ex post cost should elect a low-fixed-part and low-sharing-fraction contract and symmetrically for a firm expecting a low cost.⁹

But one of the most policy-relevant points is that the choice of a regulatory structure depends on the hierarchy of concerns. In practice, at least four groups of parties are involved: consumers, firms, the government and the regulator.¹⁰ Users worry about prices, service quality and reliability. Firms typically worry about profits, risks and market power. Governments, who are often the dominating player in the context of the reform of the sectors covered here are interested in reducing the fiscal burden imposed by the utilities and even are sometimes trying to generate a flow of resources through the reform process.¹¹ They are also concerned with the need to deal with unions as the restructuring of these sectors often results in labor redundancies. Finally they also have to deal with environmental and distributional issues. The regulatory agency is motivated by its reputation and its members' remuneration.

If the economist's concern for efficiency were shared by all, fixed price contracts would dominate all other forms of regulation. But when risks are involved, the participation constraint is likely to be binding, and social worries are at work, the best solution can be to go for some type of cost-plus regulation.

4. Prices and public service constraint

When the good or service is sold in a market, the regulatory mechanism also has to cover pricing rules. Ideally, from an economist viewpoint, they should reflect the normative principles mentioned earlier, but in practice, they end up covering the multiple concerns just like any other decision on regulatory matters prepared through a political decision making process. This is why it should come as no surprise that the best pricing policy for network industries is generally likely

⁹ In practice, this could mean for instance that the firm will accept a low guaranteed return, in exchange for a commitment by the government to take on any cost overrun for a firm expecting large ex post costs. For the firm expecting low costs, they would accept receiving no guarantee but demand to keep any profit, whatever its level.

¹⁰ The regulator is ideally independent of the government. Actually, the regulatory entity as well as regulated firms are themselves complex organization made of individuals or groups with divergent interests. For a recent survey see Estache and Martimort (1997).

¹¹ In the French water industry, three large firms supply the distribution and sanitation services to thousands of municipalities. Clearly, absent any coordination of municipalities at the national level, the operators are the dominant players.

to be again a variation on the theme of two-part tariffs. Indeed, the ideal (i.e. normative) price system would have to be discriminatory because of the trade-offs imposed by the public authority between efficiency (which commands marginal cost pricing) and all the constraints imposed by the design chosen for the regulatory mechanism. But since the government usually adds a constraint of no-price-discrimination, the only way out is to propose to all the consumers an identical menu within which they choose the cheapest tariff or menu of tariffs.

The possibility to practice non linear pricing, particularly two-part tariffs, allows the reduction of some divergence between efficiency objectives and redistribution or equity objectives, but it creates new difficulties, particularly the risk of excluding some users. First, note that two-part pricing can be a solution to the conflict between efficiency and budget balancing. First best and cost recovery are no longer antagonist if the unit variable price is set equal to the marginal cost and each consumer has to pay a fixed part equal for example to the budget deficit divided by the number of customers. But this argument is valid only if customers are homogeneous enough for the fixed part payment to exclude none of them from consumption.¹² On the contrary, with heterogeneous potential users (for instance heterogeneity in terms of revenues), the fixed part of the tariff can be too high for small users leading to the exclusion of some potential users.

The exclusion of a group of consumers raises two types of problems. On the one hand, its consequences are difficult to accept on equity grounds, particularly for services such as water or electricity. On the other hand, exclusion has negative effects even from the efficiency point of view. Actually, as long as the excluded (poor) consumers have a marginal propensity to pay higher than the marginal cost, their exclusion results in a welfare loss. Consequently, a tariff with a fixed part high enough to recoup all the fixed costs can be unacceptable. One has to arbitrate between the distortions created by an increase of the unit variable price and the negative impact of an increase in the fixed part of the tariff. Despite this caveat, two-part tariffs remain more efficient than Ramsey prices.¹³ In some cases, a combination of a two-part and linear tariffs, even if it is not a first best solution, may be welfare improving within the social constraints of universal service. For instance, in the Belgian electricity sector, a "social tariff" without any fixed part is reserved for very small consumption. This would suggest that the restructuring of a sector should result in concession areas that cover homogeneous users or that the universal service obligation attached to the concession should be financed either by public subsidization or by an acceptable cross-subsidization.¹⁴

However, when universal service concerns or redistribution concerns (between rich and poor, country and cities, firms and households...) are important to politicians but the government is trying to minimize or eliminate any contribution to the service to address fiscal concerns, there

¹² In some instance, it may not be enough to address the fiscal issue since the concession area may include homogenous consumers who are all poor and the government may end up having to subsidize the fixed part. It has however the advantage of increasing the transparency of the subsidy scheme.

¹³ Ramsey prices are the most efficient linear prices when the producer has to recoup all its costs.

¹⁴ And this is probably easier with a small "high-cost" minority.

are new trade-offs and homogenous areas may not be the preferred option anymore.¹⁵ This ends up influencing the optimal design of the restructuring (i.e. the concession area selected in the unbundling process have to allow for heterogeneous consumers) but also the design of the price. Non linear tariffs remain a good solution because they allow to generate enough revenues on the "rich" segment to perform this redistribution. This is true without any competition but it is still more pertinent in case of skimming of the best clients or bypass threats. Actually, if a public firm loses its richer clients attracted by better outside offerings, it is no longer able to cross-subsidize the poor segment and the linear price will necessarily be increased since the same fixed costs are to be paid by a smaller number of users.

From a practical point of view, for many infrastructure services like gas, water or electricity one can distinguish between a demand for access to production capacities and a demand for the use of these capacities. So these two demands should be charged separately. When a potential user wants to be connected to a network, the operator has to install a line and a meter for which the user will pay an installation charge and a periodic subscription/maintenance fee. This fee can be interpreted as the right to consume the good up to a certain pre-defined maximum amount. This part of the tariff can be dependent on the capacity subscribed but it is independent of the quantity consumed. Since the connection is the signal of a future consumption, the operator has to develop his production capacities to meet this new demand and the subscription charge should be equal to the cost of installing these extra capacities. After the connection, the quantity consumed (within the limits of the subscription) is metered and each unit should be charged at its short run marginal cost, which is essentially the cost of raw materials and primary energy used in the production of the good. Depending on the complexity of the measurement in place, the unit marginal price will be constant, or variable with the quantity consumed, or time dependent, etc. The potential outstanding issue here may be that this approach may be hard on the poor who cannot even afford the connection fee. Additional financing mechanisms must be considered to address the combination of efficiency and equity objectives.

Finally, when designing the regulatory regime, the reformers should remember that upstream vertical integration can be a device for increasing the control a firm already has on its competitors.¹⁶ By controlling the delivery of certain strategic inputs, a firm can get a dominant position on its final market. In network activities, the most sensible example is the control of access to transport and distribution infrastructures by a firm that is also a user of this infrastructure. Access to a gas pipeline or a transmission line by any generator is a key aspect of competitive energy markets. A key instrument of this aspect of competition is the definition of the proper rules of access pricing but it is beyond the scope of this paper.¹⁷

5. Information, institutions and the design of concession contracts

¹⁵ Moreover, building a large number of small homogenous concessions is not the best way to benefit from economies of scope.

¹⁶ For a survey of the problems created by competition in network industries, see Klein (1996)

¹⁷ See for example Laffont-Tirole (1994) and Armstrong, Doyle and Vickers (1996).

The specific design of a concession contract is just as strongly dependent on the ranking of government's objectives as the other aspects of the regulatory scheme. Basically, the design of the contract depends on the information concerning the good or service to provide, including a good understanding of the nature and form of demand which often changes with changes in the regulatory arrangements, in particular the pricing rules. It also depends on the information on the firm itself available to the regulator *ex ante* and *ex post*.

The richer the *ex ante* information to all parties involved, the closer to optimal planning the contract can be. The richer the *ex post* information available to the regulator, the wider the set of enforceable mechanisms that can be used to mimic the optimal planning. Consequently, the design of contracts is totally subordinated to the ability of the government to collect valuable economic information and to punish false reporting and deviations from contracted decisions. Its degree of complexity should also depend on the resources the regulation agency can devote to implement it and on the ability of the firm to understand what its interests are. Finally, the choice of the contract should also depend on the speed with which the concession decision has to be made. If the fiscal objective is pressing for instance, the government may not have the time to collect the information required to come up with an efficient and equitable contract. In some other cases, speed is important because signaling the commitment to reform matters more in the short run than the reform itself. Many of the first concessions in Latin America at the end of the 1980s and early 1990s illustrate this school of thought.

Whatever the explicit objectives of the concession may be, the contract should contain an explicit list of penalties that will be incurred in case the rules are not respected.¹⁸ In a very well tailored contract, sanctions are such that the parties do not respect the rules when and only when it is optimal not to respect them. Too high penalties give the contract a rigidity that is inefficient. Only credible threats should be included into the contract, which means threats that the regulator will have an interest in implementing if a violation occurs. Non-credible threats are bad because their presence in a contract weakens the reliability of the whole text. As the firm (rationally) does not take them into account and as the regulator (rationally) does not punish it, the agency will get a reputation of no-toughness that will induce the firm to "over violate" the rules or to try to renegotiate any decision it does not like.

When the work of the regulatory commission may lead to a decrease in the firm profits, the firm managers have an incentive to try to "capture" the commissioners. These risks of capture are very serious in developing countries. To prevent them, cost-plus is a good rewarding system since it essentially sets the rent available but as shown above, it is not a very powerful source of incentive to cut costs. Moreover, a cost-plus approach requires a monitoring of the costs which can be manipulated by collusion between the firm managers and the agency employees in monitoring. To prevent the regulatory staff from being influenced by the firm, a number of institutional features should be put in place while creating the regulatory agency. For instance, there should be restrictions on future employment of regulatory staff by regulated companies.

¹⁸ The economist tries to synthesize them by the promotion of welfare within the constraints imposed by the inconsistencies between multiple goals of the government policy.

Also, staff salaries should be high enough to minimize the incentive for corrupt behavior. This is the purpose of the "regulatory fee" that is now commonly added on to privatized utilities bills.

The second risk of capture of the commissioners is by consumers, or at least by some activist minorities of consumers. It is in fact very similar to a capture by politicians in the sense that the side payment they receive is not made in cash (like is mostly the case for capture by the firm) but some kind of social reward. To prevent a capture by a minority, it is generally recommended to ensure that the commissioners should be nominated or elected for a fixed period and protected from arbitrary removal when the regulator and the government do not see eye to eye. In fact, the optimal design of the regulatory commission is a much more complex theme than we make it appear here (staggering nominations of commissioners to ensure continuity, eligibility criteria for commissioners, how to ensure the political independence of the regulators from the political power....) but its discussion would be beyond the scope of this paper.¹⁹

From a dynamic point of view, the relation between the firm and the regulation body is a recurring one so that the whole regulatory mechanism should take account of the forecasts of both parties. The possibility to commit to specific decisions or to renegotiate the contract will induce strategic behavior on both sides. As time goes by, the regulator gets information on the behavior of the firm, on its characteristics and on potential challengers. In any information acquiring process, postponing decisions until the deadline is efficient. Consequently, from the regulator's viewpoint, the ideal contract should not commit him too much and it should not be signed for a long term. Doing so, the regulator hopes that he will be able to franchise a new firm with a new technology or to oblige the former firm to accept a new contract, more in view with its distributive concerns. But we know that the regulator receives also information from the firm itself, for instance through the ex ante revelation of its cost value. So, a short term contract can be easily interpreted by the firm as a device to extract information today in order to extract profits tomorrow. On this basis, the firm will not disclose to the regulator all the information necessary to implement the contract efficiently. To get full revelation of inside information, the firm will demand a long term contract with a commitment not to renegotiate it.²⁰

The other argument in favor of long run contracts with hard commitments is the necessity of sunk investments, that is of expenditures that the firm cannot recover if it does not stay long enough in the market. A short term contract or a long term contract easily renegotiable will systematically induce underinvestment. On the contrary, if the contract includes a reimbursement clause of the specific investments in case of breach by the regulator before the term or in case of a franchisee change at the term, the firm has an incentive to overinvest. Also one has to take account of the incompleteness of contracts. The longer the term, the more incomplete the contract is. In an industry with very intense R&D, contracts should be signed only for short periods. In a dynamic framework, the regulation device will influence technological choices and the investment

¹⁹ See Estache (1994, 1997) and Smith (1995)

²⁰ A possible alternative suggested recently by a few authors would be that contracts focus on payoffs much more than on trying to anticipate all possible physical contingencies in the contract, as discussed in Maskin and Tirole (1997).

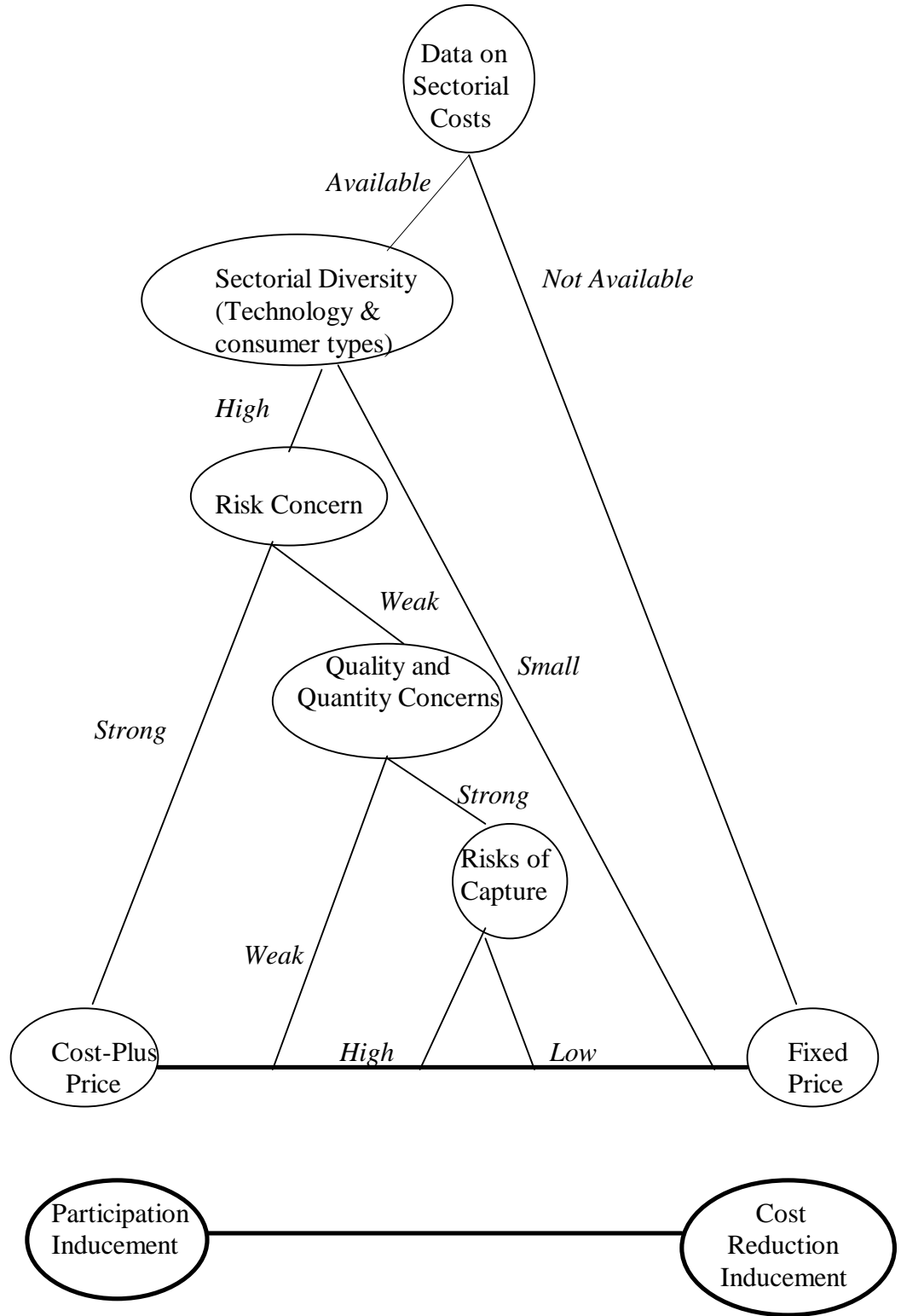
pace. The smaller the profits allowed to the firm, the smaller the amount of its investments. Moreover, if instead of trying to limit the firm profit, the regulation is aimed at limiting its rate of profit, like in the former US rate-of-return regulation framework, the firm will chose inefficiently overcapitalized production techniques.²¹

Finally note that when uncertainty becomes small, reimbursing the cost to induce efficient revelation of information becomes valueless. Only the moral hazard problem remains relevant and under risk neutrality, the contract converges to a fixed price contract. For example the dynamic evolution of contracts in the French weapons industry show that as a project evolves over time, the contract resembles more and more a fixed price contract. This may be explained by the fact that the government acquires information about the firm's cost function. In distribution concessions where the information set of both the local authority and the concessionaire change with time, the contract should also evolve from cost plus to price cap.

This discussion suggests that the reward of the franchisee through market sales and/or through subsidization will be a compromise between poorly-incentive-but-secure cost-plus rules and low-cost-inducing-but risky price-cap rules. Depending on the informational gap and on the priority of concerns, the contract will be somewhere in between these two extremes like is illustrated by the figure on next page.

²¹ This concern has to be weighted however against the possibility that without some sharing of the risks facing potential concessionaires, none will be attracted. In those cases (and depending on where risks come from), rate of return regulation allows a risk sharing with users of the service that is not as easily structured under price cap regulation.

Figure 1: **Optimal Contract Schemes**



6. Award of the concession contract

The government can extract at least a part of the knowledge owned by the potential concessionaires:

- ex ante, by organizing a bidding procedure to allocate the concession. It is a way to oblige competing candidates to reveal elements of their inside information, like their profit forecast or their average cost of providing the public service.
- ex post, by organizing a yardstick competition based for instance on ex post costs if they are publicly known.²²

The three basic methods for allocating a concession are administrative decisions, lotteries and auctions. They differ in terms of speed, transaction costs, efficiency, equity and fiscal payoffs and hence impose new choices among trade-offs. In the administrative process, a commission holds comparative hearings to decide which applicant is the most worthy. When there are many markets to concede, this "beauty contest" method breaks down under its heaviness because there can remain a big backlog of unassigned licenses. With lotteries, the method changes drastically, each market being assigned randomly among the applicants. Lotteries are very good for assigning concessions quickly but the winners are not necessarily the best candidates in terms of their ability to operate the markets, consequently they are not necessarily the ones who want to pay the highest concession fees. An auction offers two advantages over the former alternatives: it can raise revenue, which is good for the seller, but it can also be designed to identify the firms with the highest use-values for the market to be conceded, which is good for efficiency. But this again implies trade-offs which have been assessed in other contexts.²³

Auctions are just a way to organize competition in certain circumstances where there are not too many potential buyers and where the seller thinks he will be better off under this kind of "collective bargaining" than under any succession of bilateral bargaining rounds. Many exchanges are performed under auctioning. For each item, the main variables that can determine the choice among trade-offs are based on three elements that need to be determined by the auction: How is

²² To organize yardstick competition, one needs several identical activities operated by different firms (for instance local telephone in separated regions) and the differences that are observed between the performances of these firms are to be attributed to moral hazard variables, that is to the unobservable decisions of the agents. The regulator can promote efficient (while non observable) decisions within the firms making the net revenue of each firm a decreasing function of its own cost (which is an incentive to decrease it) and an increasing value of the average cost in the other firms.

²³ It has become standard to distinguish four primary types of auctions which can be used either in single object or multiple unit auctions: English and Dutch auctions which are open oral procedures on the one hand and first-price and second-price auctions which are sealed-bids procedures on the other hand. See for instance McAfee and McMillan (1987)

the winner selected? and How much and how does he/she have to pay?²⁴ On the basis of the answers to these questions, one has to give a valuation of the different institutional rules governing the exchange and these rules are important because they can affect bidding incentives and therefore the terms and efficiency of the exchange.

In practice the bidding process generally starts with some type of prequalification of potential bidders based on technical and financial criteria. This process reduces the number of bidders, which is bad for competition, but it also reduces the risks of non-compliance by unreliable bidders. It is yet another way of revealing information for the regulator. The next relevant institutional feature is the way in which the winner is selected. Here also, there are many options. Until recently it was common to rate various aspects of the technical and financial proposal and add up the results in a weighted or unweighted average. This process lacks the transparency and hence the efficiency that many potential bidders would like to see. So the two most common options in the award concession contracts boil down to this: (i) the winner is the highest payer for the right to provide the service or (ii) the winner is the bidder offering the lowest price to be paid by the consumers (in both cases for a set level of investment and quality requirements).²⁵ The first criteria favors the fiscal objectives and aims at maximizing the revenue to the government (this is the model typically analyzed by the literature on auctions). The payments are typically made as a lump sum at the beginning of the concession period or as an annual payment (which often boils down to a rental fee for the use of the existing infrastructure made available by the government). The second model focuses on the interest of users and ensure price minimization rather than revenue maximization.

These two options are typically not equivalent in net present value terms. Even if they were designed to be equivalent, their equivalence would depend on a number of unrealistic assumptions, not the least the effectiveness of the regulator in controlling the concessionaires' efforts to give the various opportunities shifts in demand and supply over time to generate short term rents. Moreover, the concession agreements tend to be dynamic, the payment agreed under the first type of decision rule is seldom renegotiated while the average tariff is often renegotiated in such a way that the concessionaire ends up with a larger share of the rent created by the efficiency gains achieved by the monopolist concessionaire. One of the advantages of the second model is that it allows faster decisions. Under the first model, the government needs to come up with a minimum payment or a reservation price if it wants to avoid the risk of having to subsidize the service. When the value of the assets involved is poorly known (as is typically the case for water companies for which most of the assets are underground), it can take quite a bit of time to

²⁴ Note that in an auction, the bids are not necessarily the price that the winner will pay to the seller. It can be any characteristics of the exchange as far as it can be measured and verified: quantity to produce, sale price, quality of service etc. Argentina's 1993 cellular license auction illustrates the variety of public policy purposes to which auctions can be dedicated. Competition was not over price but over which bidder could offer to set up cellular telephone service in the fastest time. A consortium including GTE and AT&T won by promising to provide cellular service across a vast area of Argentina's countryside in only one month. For the water concession of Buenos Aires in April 1993, the candidates had to compete by proposing a discount coefficient to multiply the computed "public tariffs".

²⁵ Williamson (1985), chapter 13

come up with a reasonable reservation price. The slowness of this process is not necessarily consistent with the need for the concession to take place quickly to make the most of the political momentum. In sum, the tradeoffs are between revenue, users' well being and timeliness.

The linkages between auction design and objectives are not the only sources of complexity. Another is the risks of cooperation between participants to the procedure. The first possibility is a collusion between buyers to limit the price to be paid by the winner. This type of misconduct is most probable in oral auctions (because the "cartel" members can watch each other) and in repeated auction (the bidders can decide who will gain each item in his turn). The second possibility is a capture of the auctioneer. He can be captured by the seller when the commissioner misreports the quality of the goods; a consumer with a true valuation less than the seller's valuation will tender too high a bid and win the auction, which results in an inefficient allocation. On the contrary, the commissioner can be captured by a (group of) buyer(s) which results in a too low sale price as compared with the predictable price in a fair auction. These different misconducts can be limited if the procedure is widely public and if the participants are numerous. Note that the collusion may be stimulated by the design of the auctions. If the government imposes an upper limit of the shares anyone bidder can own of the deal, it is very likely that various bidders will have a strong incentive to collude. In Brazil's railways concessions for instance, no single operator could own more than 20% of the shares. There were never more than 6 or 7 potential candidates to the concessions so that it was clear that they had a strong incentive to create consortia together and to figure out how to share profits after winning the bid as a single candidate. Similar issues can arise when the government is trying to impose redistributive criteria on the design of the auction by requiring the participation of minority owned or locally owned firms as part of international consortia.

The final design of an auction procedure should take account of the quantity of information the auctioneer wants to give to candidates. For a unique concession assignment, the more the candidates know about each other, the greater the chances the prize to be attributed to the highest valuation bidder. On the other hand, in a multiple-round auction, with a good knowledge of other's valuation, candidates can collude to organize alternate assignments of prizes. If there are several items to allocate, like the concessions for electricity distribution or various segments of highways, the decision to auction the concessions in sequence rather than in a large simultaneous auction depends on a delicate trade-off. On the one hand, a sequence of elementary auctions has the advantages of administrative simplicity and of instantaneous plan revision: at the end of each sale, every bidder knows exactly what everybody got and what remains to sell. But on the other hand, predatory bidding and other problems can occur. MacMillan shows that many of these drawbacks can be avoided using large simultaneous auctions or, to keep some of the advantages of sequencing, a simultaneous auction but with multiple rounds. In this case, the problem is to define clear stopping rules: the auction should last a finite time easy to compute and with aggregation possibilities until the deadline. For instance, one can decide that after a certain duration only active bidders can still offer new bids and/or that the increment to announce for winning the prize is a decreasing function of the number of active bidders.²⁶

²⁶ MacMillan (1994) and more recently Cramton (1997)

7. Lessons from the implementation of these principles in Argentina's recent privatization

Argentina is increasingly viewed as a model of what private sector involvement in infrastructure services represents for a developing country not only in terms of changes in the way the services are being delivered but also in terms of the new role of the government in these sectors. Beginning in 1989, Argentina committed itself to a wide-ranging privatization program covering utilities and transport services under the National government authority. The initial objective was fiscal: the government had decided that it could no longer afford to subsidize these services nor finance the investments required to ensure their proper operation.²⁷ But in the process the government also tried to improve the efficiency with which these services were being delivered. This entailed a major restructuring of each sector and the development of a regulatory capacity which has proven to be one of the main challenges of the reforms.

The main objective of the restructuring was to introduce competitive forces wherever possible. Competition requires multiple players and one way to increase the number of players was to unbundle services wherever possible. In electricity and gas, the vertical and horizontal separation of generation, transmission and supply had been considered a sine qua non condition to the creation of effective competition in the sector. Horizontal unbundling was the key to the restructuring of telecommunications, with the division of the public enterprise into two companies getting each roughly half of the country and with the separation of activities into three service groups, basic telephony (infrastructure and local phone services), international services (international calls, telex and data) and services in competition (national telex, national data, maritime radio). The division of the national rail company into three separate businesses, freight, intercity passenger and commuter rail--which were privatized or transferred to the provinces was instrumental to the major restructuring of this sector as well. Intercity passenger services were ultimately abandoned and freight and passenger services were themselves divided into separate lines (six for freight and seven for commuters) before offered in concession to the private sector. Argentina's main container port was also divided into its terminals which were awarded to five different concessionaires.

To maximize the benefits from competition in each sector, competitive bidding was the standard way of passing on the service to private operators, although every sector adopted a somewhat different bidding approach.²⁸ An elaborate system of weighted criteria was used to

²⁷ In transport, whatever could not be privatized was to be decentralized and it became up to the provinces to decide what to do with costly infrastructure services. The decentralization of rail and ports unfit to rapid privatization by the National government led to the abandon of many low traffic services for instance. As for water and electricity distribution, they were provincial responsibilities to begin with and not all provinces have followed the lead of the national government in getting the private sector involved so that public and private enterprises are coexisting in these two sectors.

²⁸ The terms of these concessions vary significantly: 7 year exclusive licenses for telecoms, 95 years for electricity distribution and transmission, 35 years for gas, 12 years for intercity roads, an average of about 23 years for BA. access roads, 30 years for freight rail and 10 years for commuter rail (20

select the winners for the freight rail concessions while winners for the first round of road concessions were selected simply on the amount of the fee to be paid to the government. The designers of the electricity sector introduced a sophisticated system of management period which maintained the competitive pressure on the initial winners of the distribution and transmission concessions. This approach revealed the strong concern for the sustainability of the efficiency gains of the electricity reformers, in addition to the fiscal objectives. As for water, where the bidding process was the main way of introducing competition in the sector and hence critical to the success of the reform, the procedure involved multiple stages, including a prequalification, a technical qualification focusing (some would argue excessively) on the way the bidders were proposing to meet quality and investment targets and a final stage in which the financial proposal for the lowest tariff was to be the decisive stage.²⁹

Overall, however, the design of the bidding processes clearly revealed that the fiscal concerns of the government were tainting the search for efficiency through the bidding process. The search for a minimum price for the consumers was only obvious for the water sector and the access roads to Buenos Aires. In all other cases, the award criteria was centered around some form of payment to the government (cash or debt reduction) or the minimization of subsidy requirements. As for the effectiveness of the bidding procedures in enhancing competition, the results were at best mixed: while many bidding documents were sold, several buyers of these documents ended merging so that there were seldom many bidders for each transaction (i.e. 2 and 3 for telecoms, between 1 and 5 for the electricity distribution and transmission companies, between 2 and 8 for the gas distribution and transmission companies, 4 for the water company, 1 or 2 for the freight railways).

While the unbundling and competitive bidding procedures were certainly important to introduce some of the benefits of competition in the delivery of infrastructure services in Argentina, the government was very aware of the need for continuous direct regulation of these industries since once the concessions were awarded, many of these were essentially local private monopolies. This is why each sectoral restructuring was accompanied by the creation of a sector specific regulatory agency. To some extent, the creation of every agency were supposed to recognize the main reasons for the creation of a regulatory agency: (i) the monitoring of compliance with the contractual obligations of the concessionaires (ii) consumers need to be protected from the private monopolies, (iii) the private concessionaire needs to be protected from the government who still can be tempted of interfering with the private concessionaires' management. But once more, the specific approach adopted by each sector was quite different.

While the creation and staffing of the electricity and gas regulatory agencies followed the international best practice and not yet had major problems in fulfilling their obligations, the experience of the other regulatory agencies or authorities has been much more tense. The most problematic may have been the telecoms and water regulators where there are not only staffing problems (skill mix and excessive numbers) and data problems (not enough information to assess

years for the metro)

²⁹ See Crampes and Estache (1996)

costs) but also concerns with the lack of transparency of the decision making process. As for transport regulators, who have recently been merged into a single regulatory agency, the main issue has been the lack of independence from the political power. Unfortunately, there is some concern with the fact that the division of responsibilities between the regulators and their sector ministries and secretaries is increasingly being blurred for all sectors, reducing the independence of most regulators and in the process their accountability as regulators. There is also some concern that the statutory framework under which these regulators operate do not really address the need to protect would-be competitors and granted equality of access to essential facilities to ensure that competitive pressure remains in the sector. Without this responsibility and without the tools (including access rules and pricing options), the regulatory agencies are unlikely to be filling their role in the promotion of competition in infrastructure services.

In sum, Argentina's experience demonstrates that while a tremendous transformation of the way in which infrastructure services are being delivered can be achieved quickly and fairly successfully through concessions, the long run sustainability of the short term achievements is not guaranteed. The preference for speed of reform and fiscal concerns and the nervousness about the risks of not being able to attract private investors have generally dominated the design of the concessioning processes as in water and telecoms for instance. This is why information gaps for the regulators are still quite significant for most concessions and information generation processes are still being developed sometimes over 5 years after the private sector took over. This is also one of the key reasons why the sustainability of the concession based reforms and of the regulation of private monopolists through contracts will depend on the strengthening of the regulatory agencies, on the clarification of some of their terms of references and of their accountability and on better separation of the responsibilities of the sector ministries and of the regulators....and this may be the main lesson for countries considering to follow Argentina's footsteps.

References

- Armstrong, M., C. Doyle and J. Vickers (1996), « The Access Pricing Problem: A Synthesis », *The Journal of Industrial Economics*, Volume XLIV, no. 2, June:131-150
- Crampes, C. and A. Estache (1996), « Regulating Water Concessions », *Viewpoint*, Note n°91, September, The World Bank
- Cramton, P. (1997), “The FCC Auctions: An Early Assessment”, University of Maryland, Department of Economics, Working Paper Series, No.97-04
- Estache, A. (1994), « Guidelines for the creation of provincial regulatory agencies », The World Bank, mimeo, LA1IU.
- Estache, A. (1997), "Designing Regulatory Institutions for Infrastructure- Lessons from Argentina", *Public Policy for the Private Sector*, No.10, June, The World Bank:13-16
- Estache, A. and D. Martimort (1997), “Transaction Costs Politics and Regulatory Institutions”, mimeo, The World Bank, EDIRP
- Klein, M. (1996), "Competition in Network Industries", *Policy Research*, Working Paper No 1591
- Laffont J.J. and J. Tirole (1993) « A Theory of Incentives in Procurement and Regulation », MIT Press
- Laffont J.J. and J. Tirole (1994) « Access pricing and competition », *European Economic Review*, 38:1673-1710
- McAfee, R.P. and J. McMillan (1987), « Auctions and bidding », *Journal of Economic Literature*, 25: 699-738
- MacMillan J. (1994), « Selling spectrum rights », *Journal of Economic Perspectives*, 8: 145-162.
- Martimort, D. (1992) « Multiprincipaux avec anti-sélection », *Documents de travail*, IDEI, Toulouse
- Maskin and Tirole (1997), « Unforeseen contingencies, property rights and incomplete contracts », *Document de travail*, April, IDEI, Toulouse
- Smith, W. (1995), « Establishing regulatory agencies », mimeo February, World Bank, Washington.
- Williamson, O. E. (1985), « The Economic Institutions of Capitalism », The Free Press, New York.