



INSTITUT
D'ÉCONOMIE
INDUSTRIELLE

IDEI Report # 16

Regulation

September 2010

The Regulation of Monopoly



THE REGULATION OF MONOPOLY

Bruno Jullien

Wilfried Sand-Zantman

September 2010

Acknowledgements

We thank France Telecom for their comments, but emphasize that all the views expressed in this paper are the responsibility of the authors alone and do not reflect the positions of France Telecom or their representatives. We also thank Julia Bird for translating this text from French.

1. Introduction

In market sectors dominated by large fixed costs, or subject to legal barriers, the number of active firms is limited. Consequently, every firm can have an impact on the market conditions. In such situations, the State has responded to the “market power” of certain firms by taking some control over the functioning of these markets, in both more and less interventionist forms, such as controlling the level of market concentration. While most markets fall under competition laws, only certain highly-concentrated sectors are subject to a form of regulation specific to their activity; this is the case for many network industries in particular (water, electricity, telecommunications, gas, transport...). In these different examples, even if, as often the case, just a part of the sector is a natural monopoly (the distribution of water, the transport and distribution of electricity, the communications network, the physical transport network), it is the whole sector that is subject to regulation, be it either under the direct or indirect control of the public authorities for market conditions.

This control can take various forms according to the period, the sector, and national traditions. The most common form is that of a price control for a given activity, but other systems, such as controlling the number of firms operating or the type of good produced, are possible as well.

Historically, the first forms of regulation emerged during the 19th century. Public authorities faced two types of problem among network industries that resulted in subsequent interventions. In some cases, industries that required strong, often geographic, coordination instead developed in a disorganized pattern, moving the industry away from economic efficiency. This was the case for the electricity sector, the transport sector and the gas sector (as an illustration, in 1850 there were 14 different gas companies operating in London, which prevented any economies of scale). In other cases, one firm dominated the market and threatened to abuse their dominant position, such as Bell at the beginning of the 20th century.

As the preceding examples bring to light, it is technical considerations or, most importantly, increasing returns to scale (be it in production or coordination) which drive

public authorities to intervene in markets. It is therefore natural that technological developments may allow sectors previously described as natural monopolies to evolve towards competitive markets.

At the end of the 19th and beginning of the 20th century two types of solution emerged to aid the development of network industries with economies of scale: the nationalization of different firms in the sector and their regrouping into one sole public entity; or the granting de jure or de facto of a monopoly status to a private firm, subject to some form of price control regulation. In the United States, the second of these cases emerged, most often as a compromise between the dominant firm and the authorities.

In the case of local monopolies, including gas and drinking water, franchise systems with price ceilings were predominantly used (in New York gas companies were regulated in this way from 1820). These concessions were long, typically between 20 and 50 years, to allow time for profitable investments.

In the case of telecommunications¹, regulation allowed Bell to preserve their dominant status and to avoid charges that they could have faced under the *Sherman Act*. Following the loss of their licenses at the end of the 19th century, Bell witnessed the emergence of a number of small competing operators. The company's first attempt to overcome this situation consisted of buying these competitors back in order to monopolize the market, as a number of other large dominant firms had done in different sectors. The first investigations into the abuse of a dominant position drove Bell, and AT&T with whom they had strong links, to accept regulation as a means to clog up the market to their advantage. In reality, the regulation was very light, despite the *Communication Act* of 1934 and the creation of the Federal Communication Commission (FCC), in so far as the FCC undertook no research into the way by which AT&T chose their tariffs until the 1960s.

The technological advancements of the 1950s and 1960s were necessary to bring about a change in this situation, first of all with the arrival of competition in telephone equipment and then with the development of short wave communication and the emergence of MCI.

¹ See Brock [2002] for a more detailed discussion.

This latest technological advancement convinced the authorities that competition was now possible in the sector. The break-up of AT&T in 1984 only served to confirm this development. The differences in technology between local and long distance calls together with the use of cross-subsidies between these two types of call led to the presence of regional monopolies for local calls, and competition over other types of call.

In Europe, and particularly in the United Kingdom, the adopted system quite quickly became state run companies. For both water and gas, the legislations of the 1870s (*Gas and Water Facilities Act* and then the *Public Health Act*) permitted the municipalities to take control of private firms to form these state run companies. The private firms were too small to ensure a sufficient level of investment and to prevent competition from being detrimental to the quality of service. For national network industries, the solution was the merger of companies and nationalization. In the telecommunications sector, this solution was adopted in two stages, first of all with a nationalization process for long distance calls (1896) and then for local companies (1911).

The regulation which emerged alongside the privatizations of the 1980s has therefore shaken up the traditional British management of network industries. It was the *British Telecommunication Act* of 1981 which opened the way not only to a liberalization of the sector but also to incentive-based price regulation. At the beginning of the Eighties, the dominant position of British Telecom (BT), more than any technological aspects, led to this regulation being put in place. Having previously envisaged an American system based on profit control, alternatives more favorable to economic efficiency appeared to exist. The recommendations raised in the report by S. Littlechild in 1982 served as a basis for telecom regulation. A price control was established for a certain number of services, over which BT retained a de facto or de jure monopoly, by fixing price ceilings indexed by inflation but also taking into account possible advances in the industry. The system, RPI-X, gives the firm incentives to improve their efficiency by leaving them the benefit of all the reductions in costs realized for a period specified in the regulation contract.

Following the *Telecommunication Act* of 1984, BT faced some competitors, in particular Mercury. Mercury concentrated on long distance calls, so the problem of access to the local network, over which BT still held the de facto monopoly, needed to be solved. To avoid the network effects from restricting any potential competition to BT, the regulator imposed the

complete interconnection of networks and fixed the access tariff, basing it on the costs of providing access, which gave the competing operators access to some infrastructure on BT's network.

Finally, even if eventually this regulatory scheme turned out to be very favorable towards BT, both because of the weakness of Mercury, and because technological progress ensured that BT could easily reach the productivity objectives imposed by the regulator, the English regulatory authorities introduced one of the first incentive based regulation systems. Over the last 25 years these regulation methods have spread out across all developing countries, replacing the original method of profit control.

In this report we do not aim to relate all the experiences of regulation. Instead, we explain the economic motivations behind the main methods of regulation outlined above.

The second section of this report is a brief reminder of the presiding rules as to the choices of production efficiency and the problems related to monopolistic structure.

The third section presents the principles of regulation which are applied independently of the sector considered. To start with, we return to the reasons which explain the separation between the regulator and the regulated firm. Then, we set out the traditional rules of control, that is to say the Ramsey-Boiteux pricing and rate-of-return regulation. The third part of this section discusses certain theoretical aspects of incentive based regulation, taking into account the asymmetries of information between regulators and regulated firms. Finally, we set out the alternative means of control to direct regulation.

The fourth section of the report focuses on regulatory aspects more specific to network industries, for instance the problems related to the pricing of access to a network and to interconnections. We present the rules for optimal regulation of access to a network in the framework of a monopoly possessing an essential facility, and we discuss the link between these optimal rules and the commonly applied system of the *Efficient Component Pricing Rule*. Then we discuss the problems raised by the choice of termination charges in a framework of bilateral access (« two-way access»), by highlighting the possibilities of collusion between firms controlling different competing networks.

The fifth section discusses the main principles of regulation, focusing on the difficulties involved in putting regulation into place. In particular, we set out the costs of regulation

resulting from the regulator's limited capacity to commit, and the inefficiencies which result from this limited commitment.

2. Social efficiency and monopolistic behavior

Before dealing with the problems which arise due to the presence of a monopoly, and hence from regulation, it is useful to have a normative benchmark.

In a market, and more globally in a trading relationship, efficiency should lead to the realization of all surplus generating transactions. These are trading situations in which the buyer's value of a good is higher than the production cost. Following this, social efficiency should result in the production and exchange of all units for which the marginal cost² of production is inferior or equal to the buyer's value for the good, his marginal utility. For the last unit traded, we obtain the classical efficiency rule:

$$\text{marginal cost of production} = \text{marginal utility of consumption}$$

We will see that this efficiency is reached in a competitive market, but not in the case of a monopoly situation. To take a simple example, we consider a sector for which the marginal costs of production are non-decreasing with production, and on the demand side, the value given to each unit is decreasing with the number of units sold. Of course, for each side of the market (the buyers as well as the sellers) to accept to participate in the trade, each party's surplus must be positive and then the price must be between the production cost and the consumer's value for the good.

In a competitive situation, where a single actor (a firm or consumer) has no influence on the price, the rule of efficient allocation is achieved. Each firm, facing a market price, is going to produce while the private value of the transaction is positive. From the demand side, the consumers buy the good while their value for an additional unit is greater than its price. From this reasoning we obtain at the equilibrium:

² The marginal cost is the increase in cost induced by the production of an additional unit.

marginal cost of production = price = marginal utility of consumption

In a market economy, the price is an intermediary of exchange which allows social efficiency to be reached and which determines the means by which the surplus from the exchange is shared.

What can be said about a non-competitive case? To simplify, we consider here a firm in a monopoly position.³ This means that now only this firm is going to set their price; they can of course choose a price equal to the marginal cost, and obtain a strictly positive net profit due to all the infra-marginal units for which the price is superior to the cost of production. They can, however, also decide to increase their price by a little bit, so that the demand they face decreases but that the margin on all the units sold increases. As a result there is a trade-off between the quantity sold and the margin that the monopoly can earn on each unit. This trade-off ensures that the monopoly moves away from the efficiency rule to propose a price that differs from the marginal cost.

Non-competitive situations are common in industries where large investments are needed for production. In such cases there is little or no competition, which can result in situations where the dominant position is abused. Then, even when putting aside any regulatory or informational issues, it is difficult to force a monopoly firm to choose the efficient price rule, since pricing at the marginal cost could prevent the firm from making a return on their investments. The following section therefore focuses on the best way to take all of these different problems into account.

3. Theory of Regulation

3.1. Separating the firm and the regulator

³ In our simple framework of non-decreasing marginal costs, this monopoly position can be explained by the need for large initial investments, such as the construction of the network.

The majority of network industries have a regulation which works due to the separation between the activities of the planner-regulator and the manager. This separation, however, has not always been in place, particularly in Western Europe. From the post-war years until the 1970s, there was often such a large confusion between these two activities that it is necessary to return here to the initial arguments for separation. More precisely, there are two aspects to analyze: firstly, the separation of the regulator's activity from production, whether or not this separation be associated with a form of privatization; secondly, the movement from a monopolistic structure to a competitive structure, over a part or all of the sector's activities.⁴

It is difficult to understand the separation of activities without considering the problems that characterize the production of collective services. These are in three forms: firstly, it is difficult to conciliate diverging interests and objectives; secondly, it is difficult to obtain all the relevant information, which is spread out and used in a strategic matter by those who hold it; finally, the services require a capacity to commit in a credible way on the choices which will be made in the future, however this capacity to commit is essentially limited.

In the combined case, the regulator and the team directing the firm are the same body. The problems of asymmetric information can therefore be greatly reduced, guaranteeing, for a given level of technology, a supply offer best reflecting social preferences (in terms of allocative efficiency). However there are two problems associated with this type of organization. Firstly, choices are not transparent. The directors, close to the production activity, are going to take the interests of the employees into account, and are likely to place less of an emphasis on social well-being (the satisfaction of the rest of society). Similarly, it is difficult within this framework to use incentive-based systems that would allow costs to be reduced (or productivity increased), thereby compromising productive efficiency, defined as

⁴ We note that the regulator also has the choice of market structure at his disposition. In addition, the development of competition, even if it sometimes arises from mistrust towards monopolistic structures, is also a form of regulation which can be called upon when the classical instruments cannot be used.

the choice of the most efficient technology of production. In fact, even if the firm's results are not satisfactory, the directors (regulators and managers) cannot easily punish the firm or limit the provision of subsidies.

When management and regulation are combined, therefore, there is at best allocative efficiency, but only very rarely productive efficiency, since the incentivizing tools needed to reduce costs or increase production are not available.

The problems are reversed in the separated case, whether under public or private organization. The separation of production and regulation activity allows for increased transparency and induces a clearer separation of objectives, the regulator taking charge of social welfare and the directors taking charge of the firm's interests. However, this also results in an informational barrier between the regulators and the directors. Under this framework, it is harder for the regulator to take the best decisions in terms of price and supply as it is harder for them to access the precise characteristics of the firm. Therefore, in order to limit inefficiencies in the production process (the internal rents for the firm), the regulator accepts a reduced level of allocative efficiency. In contrast, by being distant from the producer, the regulator is more independent and credible as the giver of orders. This increase in credibility in the durability of reward systems and penalties incites the directors to assure greater efforts.

This trade-off – highlighted by K. Schmidt [1996] – can favor the integrated management of service provision when both technological progress is limited and the firm and society have converging objectives. When technology is rapidly changing, or when there is a divergence between the social preference and the interests of the firm, it can be preferable to separate the production process and the regulator's activities

As we will discuss later, this separation of activities can also accompany increased competition.

Finally, therefore, technological progress and the changes in perception on the nature of an integrated firm's objectives have led to this change in the way large industrial sectors are organized. Now we need to understand the main principles that underlie the current methods of regulation.

3.2. Classic rules of regulation

Monopoly price regulation can be done in two ways: through profit control, or through price control. These two methods have implications which depend on the precise way in which they are put in place. However, as a first approximation, profit control does not ensure that firms optimize their production technology (under-investment, bad management of costs) while price controls may leave too great (or too little) a profit for firms.

Historically, rate-of-return regulation has been given the most attention. It disciplines the pricing behavior of a monopoly by imposing constraints on their profit or on the return to their invested capital. In the latter case, the authorities propose a rate of return which is high enough to attract investors but smaller than that resulting from unregulated monopoly behavior. This method is frequently used in the United-States, in particular under the form of a regulated return to capital invested. However, as Averch and Johnson showed in the 1960s, this induces inefficiencies in firms' technology choices. By over-investing in capital, a monopoly can increase their prices and approach the prices of an unregulated monopoly, while continually respecting the regulator's rule. In other words, rate-of-return regulation can lead to inefficient choices and useless investments.

It is also possible to regulate the firm by directly controlling their prices. As previously mentioned, pricing at the marginal cost does not allow monopolies in sectors with high fixed costs to secure a return high enough so that investors have incentives to operate in the sector. It would in theory be possible to completely compensate the fixed costs borne by firms, but that would involve the transfer of large amounts of public funds, which is both politically and financially difficult. We can however try to determine the optimal price to guarantee a revenue to a firm by covering their fixed costs. Marcel Boiteux (1956) produced the economic structure necessary to solve this problem. Let us suppose that the regulator decides to control the monopoly price on many markets, and that the total cost function includes large fixed costs. Pricing at the marginal cost will not allow the firm to obtain a normal profit, so the regulator must allow the firm to increase their prices above the marginal cost. In order to choose which markets should bear the most of this price rise to pay for the fixed costs, the Ramsey-Boiteux method analyzes price elasticities across the

different markets, increasing the price the most in the markets with lower elasticities. Eventually, prices are not equal to marginal costs but are closer to a mean cost obtained by attributing the fixed costs to the different services at a rate inversely proportional to their elasticity of demand.

More globally, the classical regulation methods can be regrouped into two large families: *cost-plus* and *price-cap*. In the first case, by analyzing the costs and setting a rate of normal profit, the final price is determined relative to the level of costs realized, giving an acceptable margin to the firm. In the second case, the regulator chooses a price, given as a function of an estimation of the evolution of costs, that the firm cannot adjust during a certain period of time. Even if they appear to be different in theory, these two approaches are in practice quite similar. In fact, if a *price-cap* contract lasts only one year and the *price-cap* is regularly reevaluated, it is equivalent to a *cost-plus* contract. In contrast, if the duration of a *cost-plus* contract is long, the firm can make efforts to reduce their costs and recuperate the fruit of their efforts, thereby giving them an incentive to increase their efficiency.

All the methods set out above involve transparent relations between the regulator and the monopoly. In particular, these methods of regulation rely on the principle that the costs and behavior of the monopoly are completely observable by the regulators. In reality, however, even if some elements are public and known by the regulators, it is difficult to believe this hypothesis fully. The following section therefore examines the consequences of imperfect information on the incentives in regulation.

3.3. Incentive-based regulation

Here we set out the classical trade-offs in regulation in a framework of asymmetric information. The rules of regulation given above suppose that the regulator has access to all the relevant information and from this information can easily control the behavior of the regulated firm. In reality, it is rare that this assumption of perfect control holds, in particular since the costs that serve a basis to the regulation are not perfectly observed by the regulator.

Incentive-based regulation was analyzed by economists in the 1980's, principally following the articles by Baron-Myerson (1982) and Laffont-Tirole (1986). The fundamental assumption is that the regulator does not know the marginal cost of the firm which he must regulate, and therefore cannot easily provide strong incentives for them. He has, of course, some information on these costs, but cannot say for certain if, in a simple case, the firm is efficient or not. It is also necessary to take the weight the regulator places on the profit of the firm relative to the consumers' surplus into account. In reality, the regulator has a tendency to favor the consumers over the firm, which leads him to reduce the profit of the latter by as much as possible. The regulator's objective, faced with a firm whose technical characteristics are not fully known, is to choose a regulating price and eventually a transfer which will maximize social welfare. This price must be sufficient so as to ensure the firm's participation, whatever the costs they face, without leaving them with too much profit.

If the regulator was able to observe the firm's costs, he would choose the Ramsey price associated with each type of firm. As this is not possible, the regulator must instead offer a «menu of contracts» among which the firm will choose according to their real costs. The problem for the regulator is to ensure the participation of all types of firms, so that a less efficient firm does not leave the market, but also to propose a price corresponding to the firm's costs so as to avoid an efficient firm from taking too large an advantage. If we limit ourselves here to two types of firm (a good firm with low costs and a bad firm with high costs), we can show that the optimal scheme consists of proposing the classical Ramsey price to the good firm whereas a price above the corresponding Ramsey price should be proposed to the bad firm. By increasing the price that a bad firm can propose, the demand that a bad firm would face is reduced, and a good firm is not tempted to choose this price. This distortion of course has a cost, because it produces some inefficiency for the bad firm by limiting the number of trades realized. If, however, the regulator wants to limit the excess profits left to a good firm, he must deter the good firm from choosing the tariff proposed to a bad firm. There is therefore a trade-off between efficiency and the reduction of firm profits.

Instead of assuming that the costs are unobservable, which is not always realistic, we can alternatively look at the case where the final cost of production is public but it is influenced both by an intrinsic cost parameter as above and by the effort influenced by the

directors of the regulated firm. This is the approach followed by Laffont-Tirole (1986, 1993) in their different studies on monopoly regulation under incomplete information. Fundamentally, the central trade-off between the search for efficiency and the reduction of profits remains under this modification. As the regulator wants to limit the profits left to the efficient firm, he reduces the production levels proposed to inefficient firms. Therefore, even if the most efficient firms try to keep a rent, this rent is reduced compared to the situation where the regulator pays for complete information, since the firm may always have the temptation to pretend to be less efficient,.

The analysis of Laffont-Tirole allows also for the reinterpretation of the theoretical results on optimal regulation in the more familiar framework of the choice between *cost-plus* and *price-cap* contracts. In fact, everything happens as if the regulator was proposing to the firm different possible contracts, with each contract indexed by the costs reimbursed by the regulator and the share of the revenue kept by the firm. The menu of optimal contracts is such that, the more efficient a firm is, the greater the share of their profits that come from the consumers. At the limit, the most efficient firm would choose a contract such that all their revenues come directly from consumer sales while the least efficient firm is entirely compensated by the state, who receives all the takings.

3.4. Deregulation and opening to competition

As we will develop further on, the regulation of monopolies has its limits, which results in competition having a role in disciplining market power and providing a motor for innovation in the activity areas where it is supportable.

There are few activities which are natural monopolies in the long term. The examples of electricity or the postal service or telecommunications show that technological advances (in the last example) or a better definition and differentiation of the different goods provided (in the first examples) allow, after some time, monopoly situations to progress to situations where competition can exist. It seems therefore that there is no longer any need for a sector regulator, and only a competition authority is necessary to ensure that the fundamental principals are respected (controlling agreements, repression of the abuse of market power). In sectors such as telecommunications, electricity or transport, this has led to a sort of

disintegration of activities (at least functionally) which has allowed for the introduction of competition among them.

Of course, it is difficult to put full competition in place across the whole sector, and often only certain parts are under a form of competition.

The arguments in favor of the development of competition are the same in the case of large industrial activities that were once completely in public management, and in those activities with more traditional production. If the regulator had perfect access to information, in addition to a strong commitment power, he would not need to resort to complex regulation or competition, and the firm would be naturally constrained in their choices since they are fully controllable by the regulator. The reality is different and, in a perspective of regulation, competition plays two principal roles. It provides an incentive for each firm to put the most efficient production techniques into work, and to offer the most performing services, well adapted to demand. Here again, in an evolving world, it is difficult for the regulator, and even for the firms in the sector, to know at every instant the true technology frontier. Setting up a form of competition encourages innovative techniques to emerge and makes the firms' nature and performances public. In a regulatory framework, that is to say if competition remains limited, this information could also be used to create a regulation by comparison, or yardstick regulation. We note finally that in certain cases, as a complement to competition, the opening of capital to the financial markets plays a useful role in the creation of information. The presence of financial opportunities for certain investors drives them to uncover information on the performances of firms listed on the stock market. This information is progressively incorporated into the prices and the regulator can use them to adjust their regulatory schemes (see A. Faure-Grimaud (2002)).

Apart from the gains from competition in the market, we must also note that the recourse to certain forms of competition can also be useful in situations requiring regulation. In this way, if the *laissez-faire* is not an acceptable situation in the long term, when no entry is possible and the position of the present firm is not contestable, there exist ways to avoid abuse related to a dominant position without turning back towards complex regulatory schemes.

The first of these solutions, when competition cannot be organized in the market, involves organizing a competition for the market. This is an idea that has been used for a longtime by franchises, or concession contracts, in the municipal management of public services. To determine the price and quality of service, instead of putting into place complex price-setting policies, the contractor can « take to auction » the monopoly position. This could involve, for example, making a call for offers and choosing the firm who is able to propose the lowest price. The advantage of this procedure is that the price necessarily corresponds to the characteristic techniques of the sector and is not influenced by the bias (or potential preferences) of the regulator. Unfortunately, however, there is not always a sufficient number of firms capable of providing the service and therefore to compete over this call for offers (see Dana and Spier (1989)).

Another solution to avoid these heavy procedures of regulation consists of calibrating the price or the objectives of increasing technical efficiency against comparable situations, either in other regions or other countries.⁵ These procedures known as benchmarking have the advantage of putting into indirect competition different operators carrying out their activities in potentially very different situations. This also allows for the demands of the regulator to be objectified in relation to the firm he must control. The difficulty rests of course in the heterogeneity of the particular situations (how can we compare the markets and the costs of production in two regions with very different geophysical characteristics?) and the necessity of taking into account local parameters in the determination of objectives.

These considerations form the base of the theory of regulation. An optimal price rule under complete information should lead to the regulator imposing prices equal to (or near) the marginal cost. However the presence of fixed costs and asymmetric information should lead to a movement away from these principles, orienting regulation towards higher prices and decreasing the activity of the least efficient firms.

⁵ In the manner of external reference price practices used in some countries to fix the price of medicine.

4. Applications

Each industry has its own specificities and regulation must therefore be adapted to the individual context. Consider for example the telecommunications industry. For the last fifteen years, it is the problem of access, be it one-way or reciprocal, which has been at the heart of their problems. In the case of one-way access, the main risk is that the infrastructure or facility manager may use their monopoly position over the infrastructure to exclude their competitors from the market. In the case of reciprocal access between two networks both using the infrastructure of the other, the risks are principally those of collusion and anticompetitive behaviors. These two problems and the debates on the related regulatory solutions are the object of this section.

4.1. Regulation of “One Way access”

Here we discuss the problems of regulation in the framework of a firm possessing an essential facility necessary for competition in the final market. In the case of telecommunications, this concerns the access to the local loop; in the case of gas, the access to the distribution network; while in the transport sector, it is the lines and the stations which we are concerned about. In some cases, the regulator may also influence the conditions of access to monopoly controlled infrastructure.⁶

The access charges should conciliate the different objectives: to efficiently use the network, to encourage the network owners to maintain it at the best cost, and to ensure sufficient competition on the final market without the regulatory costs being too great.

Ideally, the access tariff should be equal to the marginal cost of access, but the presence of fixed costs leads often, following the ideas of M. Boiteux (1956), to an increase in the price above the marginal cost. More specifically, the regulator must take into account the

⁶ In this way regulators have played an active role in the development of MVNOs, operators of alternative virtual mobile services using the networks of operators already in place.

fact that the network owner and the firm which wants access to this network are competing. As a result of this, the access tariff, which will be reflected in each competitor's final price to the consumer, is going to have an impact on the choices of the consumer to demand from one firm rather than another. For example, by lowering the access tariff, the regulator intensifies the competition between the two firms, therefore limiting the capacity that the network owner has to recover their fixed costs, but increasing the total demand.

In other words, the optimal access price should be based on

- The marginal costs,
- The fixed costs,
- The elasticity of demand,
- The degree of competition between the network owning firm and the other firms.

In this way, the greater the costs, the more elastic the total demand, or the greater the substitutability between the goods the firms produce, the greater the regulated access prices.

Even if the economic theory suggests choosing a Ramsey-Boiteux based price, it is an alternative rule known as the Efficient Component Pricing Rule (ECPR) which first emerged out of the discussions between economists and regulators. This simple rule, the initiative of the economists W. Baumol and R. Willig, established a link between the wholesale price and the retail price. The ECPR imposes that the access tariff proposed by the network owning firm should not be greater than their opportunity cost of competition. Specifically, if the price on the competitive market is p and the cost on this part of the market is c , the access tariff a should be such that $a \leq p - c$.

With this rule (that is taking $a = p - c$), an entering firm pays the network owner a price which compensates fully for their loss of clients. In addition, this rule guarantees that only firms of a certain efficiency will enter the market. In fact, the entering firm can only profitably sell at price p if their cost are less than $p - a = c$, therefore less than the costs of the network owning firm. Following this, with the ECPR access is guaranteed to efficient firms

and only to efficient firms. In addition, the profit of the network owner is preserved, which guarantees their capacity to finance the network.

It should be noted that the ECPR is a partial rule that was initially proposed for cases in which the retail price is controlled. If the network owner can choose this retail price, and then propose the monopoly price to the consumer, the imposition of an ECPR based access tariff allows them to keep their monopoly profit whatever the sales of the other firms.

In addition, if the prices proposed by the two firms differ slightly, or if these firms have some market power, the ECPR is not a good regulatory rule. For example, if the network owner's demand is captive, the entrant must be able to propose a lower price to capture some of the market, and therefore the optimal access tariff is less than the opportunity cost. The ECPR rule is a simple rule of regulation, which explains in part its initial success, which can limit entry into the sector only to those firms that are at least as efficient as the incumbent. This rule, however, is only optimal in simple cases (perfect symmetry between firms, no market power) and it must be amended to solve all the tasks assigned to an access tariff.

These rules also ignore the problems of asymmetric information mentioned above. In fact, access regulation has been one of the fields of action for incentive-based regulation and is discussed in the reference work of J.J. Laffont and J. Tirole (1993). One discussion put forth for the telecommunications sector is provided in their coverage of 2001. The access tariff is often determined using a *price-cap* contract. The *price-cap* can be defined for an individual product, but can also be defined on a basket of products ("*global price-cap*"). An important question is therefore how to define a *price-cap*, in particular the reference costs to be used (accountable costs or costs borne in the long-term, taking into account risk, treatment of sunk costs).

4.2. Regulation of « Bilateral access »

In the preceding analysis, one of the parties had the possession of an essential facility. It is also possible that each party needs to use the infrastructure owned by the other in order to operate on their market. This is of course the case in mobile telecommunications in order to gain access to the final consumer, or also for international calls. It is also the case in other sectors such as banking (in the access to cash distributors or credit cards). In the case of

telecoms, the questions asked to the regulator are numerous. Who should pay for the access, the emitter or the receiver? Should firms pay for access or should they have a system of « Bill-and-Keep »? Should the access tariffs be freely negotiated between operators, or regulated?

At this stage, it is useful to make some observations. First of all, in a framework of imperfect competition where the firms have made investments to build their network, the regulatory rule should allow these firms to cover their costs. In other words, at the industry level, the price must be adjusted (increased or decreased in the case of monopolistic behavior) so that the margins accumulated compensate the fixed costs. We note next that if the firms are identical and the calls of the consumers are balanced (within and across the networks), the choice of whether to charge an access tariff still has an impact. In fact, even if the revenues and the gains generated by the access tariff balance, the perceived marginal cost of each call increases with the access charge and hence the price (and the demand) is affected. In this way, even in a symmetrical situation, all access tariffs should not be equivalent and the system of Bill-and-Keep has no reason to be optimal or chosen by the firms involved.

The economic analysis of access tariffs in this situation owes much to the articles of M. Armstrong (1998) and of J.J. Laffont, P. Rey and J. Tirole (1998a, 1998b). It is recognized that to leave each firm to choose their own access tariff leads to excessive prices. The debate has therefore focused on the choice of the optimal reciprocal access tariff. When there is competition between the two networks, the authors have shown that the firms' choices can be an instrument of tacit collusion. In fact, even if the net revenues due to the interconnection are zero in the case of perfect symmetry, the impact of the access tariff on the price allows the firms to gain monopoly profits for the industry.

Therefore, in the case of competition in linear prices, the firms' profits are affected by the choice of the access tariff and the regulation imposed.

If the firms now propose some non-linear prices (with a fixed price and a variable price) to the final consumers, the net profit level at equilibrium no longer clearly depends on the choice of access tariff. If this tariff is large, the firms are going to compete over their fixed prices to attract the clients and cancel this out with additional revenues coming from the higher price of calls.

B. Jullien and P. Rey discussed these questions in detail in their reports of 2006 and 2008 on the termination of calls in the mobile industry. While this debate has largely focused on the risk of excessive cooperation between competing firms, with a high access tariff there is also a risk of insufficient cooperation as well as a risk of the eviction of small entrants to the market. These risks appear in particular in the case of discrimination between calls on-net and calls off-net. From the literature on two-sided markets, we can conclude that orientation towards costs is not a-priori optimal and that we should take into account the way in which the access tariffs impact on the final prices, the participation of different actors and the use of the service. A change in the access tariff can induce some redistribution between actors (small or large consumers, callers or receivers, etc) which needs to be evaluated. For these reasons, the optimal level of the access tariff can depend on the local conditions.

5. The limits and costs of regulation

The regulation of industries under a natural monopoly seems clear at first glance, but it cannot be put into action without producing problems. As we have seen above, the regulator is often ill-informed on the real conditions of the market (the demand or the costs) and the tariff choice is affected by this informational deficiency. In simple terms, regulation, principally in its incentive-based aspects, is a dynamic problem and the regulator's capacity to commit is crucial to ensure economic efficiency. In this section we study this point and more globally the debates related to the opportunistic behavior of regulators. This section will also be the occasion to discuss the dynamic effects of regulation. In fact, the presence of a monopoly coupled with the setting of prices has effects on the potential for entry and innovation in the sector. Finally, we discuss the potential weaknesses of regulators and their effect on regulation.

5.1. Dynamic analysis of regulation

Regulation aims to control the potentially abusive prices chosen by a firm in a monopoly position on a market. It must also be able to accompany technological progress and incite firms to optimize their methods of production. These two different objectives combined by proposing an evolutionary path of the regulatory framework through which the regulators work. Following this, in the case of the regulation put in place by S. Littlechild in Great Britain, the tariff evolution schemes were fixed for a number of years, to secure the firms' investments. This medium term commitment capacity is fundamental to how well regulation works. It ensures that the firms make good choices both at the moment when regulation is put in place, and also throughout the whole regulatory period.

If we return to the framework developed in the section on incentive-based regulation, and we extend the analysis by a number of periods, the importance of the commitment capacity becomes quickly apparent. At the first period, the regulator proposes to the firm to choose between different regulatory schemes. Through their choice, the firm reveals the nature of their costs and therefore its capacity to produce at a low cost. If, at the beginning of the first period, the regulator engages on a regulatory scheme that is stable for several periods, the firm will choose their actions in a way which conforms to the wishes of the regulator. In this way, an efficient firm will accept to produce a large quantity in return for a substantial remuneration. If, in contrast, there is limited confidence in the regulator's commitment capacity, then the most efficient firms will hesitate to make a choice which will put them at the mercy of a change in the rules in a future period. Similarly, the firm will hesitate to invest to increase their efficiency. In other words, if the regulator cannot commit not to expropriate the firms once all the relevant information (or relevant actions) has been revealed, there is no chance that the regulation will induce efficient choices on the part of the firm in the first place.

If we are interested by the dynamic properties of regulatory schemes effectively put into place, it is quite natural to focus on the importance of incentives. Take for example a system of a price ceiling, of type RPI-X, running over a period of 3 years. The idea of this

scheme is, in guaranteeing the evolution of the price over many years, to incite the firm to invest to reduce their costs and increase their margin in doing so.

Two potential problems are put forward which could put this process at risk. First of all, it is important that the scheme proposed *ex ante* is not modified during the period. It is possible that during the period of regulation, unexpected events lead the firm to make more or less profit than the scheme expected. It can therefore be tempting for the regulator to increase the tariff so that the consumers benefit from the progress realized by the firm, or reduce it if the firm encounters difficulties (like the problems of the *Soft Budget Constraint* evoked by J. Kornai (1986) on the control of firms in planned economies). In these two cases, the modification of the rules of the games in the middle of the period and, above all, the expectation by the firm of such behavior, removes the incentivizing power of the scheme. When their efforts will no longer be compensated or when the absence of their efforts will no longer be punished, the firm no longer has an interest to follow the incentives given by the regulator and any efficiency progress will be very limited.

If, however, even the regulator would have respected the regulation contract for the whole period, it is when a new regulatory contract is negotiated that the problem of expropriation can arise. Between two periods of regulation, the negotiation takes as a reference the level of costs reached at this moment. If the firm has shown their capacity to reduce these, the regulator as a consequence has an interest in demanding even more constraining efforts for the periods that follow. Anticipating this “ratchet effect”, the firm has less interest in making efforts to reduce their costs during the first period (see J.J. Laffont and J. Tirole (1993)).

These two examples illustrate nicely the trade-off between efficiency (here a decrease in costs) and the limitation of profits left to the regulated firm. At each moment, the regulator is tempted to recuperate for the consumers a large part of the gains in efficiency realized by the firm. This temptation, which exists in a static framework, is even greater in a dynamic setting and is susceptible to put all the improvements in performance of the sector back at risk. Therefore the regulator’s capacity to commit and the prevention of politicians from legally intervening are necessary conditions for regulation to function well. History has

shown, for example in Great Britain in 1997,⁷ that even in countries with an independent regulatory system and credible judiciary authorities, nothing is ever certain for regulated firms.

5.2. Regulation and market development

As shown by the reaction of Bell, who at the beginning of the Twentieth century accepted regulation to avoid future competition, regulation, be it through the market forms it induces or the resultant prices, has a long-term impact on the development of a market. This impact concerns both the possibilities of entry into the regulated part of the market and also the effects on the adjacent unregulated markets. Finally, we examine the link between regulation and innovation.

First of all, regulatory price-setting has an impact on the possibilities of entry into the market. Even if no law gives an exclusive license to the regulated firm, the regulator's choice can favor or prevent the development of a market. In this way, the choice of a low access tariff for infrastructure can favor the entry and development of a high quality infrastructure service, while a high access tariff can favor the emergence of competing infrastructures.

One of the difficulties sometimes met occurs if the incumbent firm has already paid off some of their investments, and can therefore support prices lower than their long term marginal cost, while these prices prevent the emergence of alternative suppliers. The regulated retail price, set too low, can therefore cause a phenomenon called « *price squeeze* ». The debates on the regulated tariffs in the French electronic sector are an illustration of this phenomenon.

Even if the regulated prices correspond to the long term marginal cost, in industries with large economies of scale, it is difficult for an entering firm with a weak installed customer base to propose competitive prices (Fumagalli and Motta (2010) develop this

⁷ The highly incentivizing regulation of the former monopoly based sectors, gas, electricity, telecoms and transport, resulted in very high profits in the middle of the 1990s. Following the general elections of 1997, the British government imposed an exceptional tax, the « windfall tax », against the previous commitments of the regulators, but conforming to the promises of the Labour Party campaign.

point in a framework of predatory pricing). In this way, the imposition by the regulator of tariffs near to the long term marginal cost of the dominant firm can lead to a prolongation in the monopolistic characteristics of the market.

In important cases, such as in telecommunications or postal services for example, regulated firms, often the former integrated monopolies, are only controlled over a part of their activities. Highlighted by J. Chaaban (2008), a difficulty exists in the way the common costs of different activities (regulated or not) are affected by each of these activities. In fact, in a number of cases the cost reductions used for the regulation of prices in regulated activities depend on how the activities are divided, for which there is no clear established rule. A rule that sets the tariff in the regulated sector as low increases the costs to be covered in the competitive part of the market, and can put the firm into difficulty. If, on the contrary, the rule induces a high regulated tariff, the regulated firm can use the captive part of their activity to generate financial resources and practice predatory pricing on the competitive part of the market without putting into peril their financial equilibrium. Therefore the right balance must be found.

We note finally that regulation is sometimes seen as a brake on innovation. In a study published in 1997, J. Hausman *et al.* calculated that delays in the introduction of mobile phones to the United States due to the hesitation of regulators corresponded to a social loss equivalent to at least 24 billion US Dollars per year.

With the present prices on existing technology low, the regulator can dissuade the entry of new competitors with more costly services in the short term but profitable services in the long term. However the regulator also dissuades the regulated firm from researching innovative products. The process of innovation can only occur if the firms (in place or potential) have the perspectives of greater profits than with those products (or technologies) currently in place. In terms of innovation, the problems discussed previously, related to the incapacity of the regulator to commit on the level of future prices, are particularly important when the products are not known before the development phases. Price regulation or profit regulation therefore has the effect of reducing the incentives to create new ideas or adopt new inventions.

In this way, regulation if it is sometimes necessary, acts as a stabilizer of the market. When the market is in a stage of rapid technological development, it is often preferable to adopt a light regulation, of the antitrust type, and to leave the competitive processes to develop.

5.3. The limits of regulators

The last main theme on the limits of regulation takes the regulators' own limits.

The first one comes from the limits in the capacity of any administrative entity to acquire and treat the necessary information to make an efficient decision. As a result, the development of a specific expertise is one of the arguments often put forth for the creation of specialized regulatory authorities.

In addition, in classical economic analysis of regulation, the regulator and the social planner have the same objectives. In reality, however, regulators are agents as any other who can have their own preferences, be subject to influence or simply affected by taking their own personal interests into account. Therefore we need to ask "who regulates the regulator?" Different contributions have analyzed these elements to understand the limits of regulation.

The classical "capture" analysis of J. J. Laffont and J. Tirole (1991) focuses on the regulatory costs due to the possibility of a tacit agreement between the regulator and the regulated firm. Let us suppose that the regulator is in charge of gathering information on the firm and transmitting this information to a superior authority. If the regulator obtains precise information on the costs, the firm can lose all their informational advantage and can consequently lose all of their rents. They therefore have an interest in convincing the regulator to report information in favor of the firm, which the regulator may be prepared to do, knowing that the two parties are committed to a long term relationship. We note here that, in a framework of asymmetric information, the stronger the incentives in the regulatory scheme, the greater is the temptation to collude and hence it is necessary to remunerate the regulator well enough to ensure that he makes every effort necessary to collect the information. This is why, in order to get around the problem, it can be optimal to reduce the temptation to collude by reducing the incentive power of the regulatory scheme.

In other words, the potential presence of collusion leads to regulatory rules that are much less rigid and reactive to the informational available.

In a more general way, a regulator having such powers can induce a movement of resources from private actors to inefficient lobbying activities, at the detriment of productive investments (innovation, quality, etc)

The potential link between the regulator and the industries which he is in charge of regulating is therefore to be wondered. This topic is all the more discussed since some regulators are chosen among industrialists and reciprocally certain regulators rejoin the industry at the end of their mandate. This type of coming and going should not be condemned in a systematic way. In fact, as the regulators are often too short-termist in their decisions (in particular in the rules of allocation and the profitability of investments), authorizing this type of transfer leads them to take the interests of the firm further into account. Following the analysis of D. Salant (1995), this policy of « Revolving Doors » can be optimal as the risk that myopic regulators engender underinvestment is stronger than the risk that collusion occurs between the regulator and the regulated firm.

J.J. Laffont (1996) also highlighted the importance of electoral games when analyzing regulation. In fact, the preferences and ideological choices of the political leaders, themselves chosen by an electoral process, could be seen as one of the greatest constraints to regulation. As a result of the electoral process, the weights given by the regulator to the firm's profit can be either larger or smaller according to the political majority, which would in turn lead to more or less incentivizing regulatory schemes, and also produce a risk to the community. By having minimal regulatory rules, that is to say rules that hold firm over time, the community would be protected from politically originated regulatory risk, therefore maybe increasing their well-being. Of course, the political risk must be sufficiently large relative to the possible variations in the fundamentals (distribution of costs, information on the demand) in order that simple rigid regulatory rules are preferred to a more reactive regulation allowing fast adjustment to changes in fundamentals.

This analysis was completed by A. Faure-Grimaud and D. Martimort (2003), who combined both the changes in political preferences and the risks of collusion between the regulator and the regulated firm. In general, they defend the concept of choosing a

regulatory system which is independent to changes in the political majority, thereby increasing the theoretical risks of collusion. In addition, the presence of the same regulator facing several successive political representatives with short mandates and different preferences can result in a ruling policy that overly constrains the set of possible measures to limit the consequences of political fluctuations in the future.

Market developments make regulation in practice necessarily complex and continually changing, and therefore it is susceptible to the effects of exterior influences. Sometimes it is the lack of reactivity on the part of the regulator which is deplored. At other times, it is the too large a dependence on current events or the pressures of short term objectives. Therefore, finding the perfect balance between rule and discretion remains one of the greatest challenges in the principles governing the practice of regulation.

6. Conclusion

As all interventions on markets, price regulation remains a delicate exercise. If it is too relaxed, the firm may exercise their market power at the expense of social efficiency. But if it is too intrusive, it replaces market mechanisms by allowing the regulator to bend the industrial or commercial decisions of the firms concerned. Finding a suitable degree of intervention is inherently difficult in any action trying to control the behavior of private economics agents. If the fear of excessive interventionism interrupts the innovation process, or limits the entry of potential competitors, it can result in private decisions which do not reflect the long term interest of society. Along these lines, I. Segal and M. Whinston (2007) have recently highlighted that in the case of innovative industries, competition policy must come as a compromise between the protection of the innovator's profits, which compensate their innovation, and the limitation of anticompetitive behavior.

When we consider the regulation of a given industry, it is therefore not only the form of the intervention that must be evaluated, but also its coverage. For instance, the coverage of price regulation in the telecommunications sector has considerably reduced over the last

two decades. This includes both the introduction of competition into numerous market segments (long distance calls, mobile technology, data, and directory service) but also the deregulation of the retail price and a regulation focused on wholesale markets. The European Commission's intervention on the market for roaming shows, however, that in some situations, the authorities may believe that a return towards regulation is necessary.

To conclude, we emphasize that choices concerning the coverage and timing in the development of regulation can have important consequences. The telecommunications sector, as well as other regulated sectors such as transports or energy, is currently the subject of numerous developments which require important investments in a context of great uncertainty. In this critical phase, the regulator must be attentive not only to ensure that the firms in place do not hinder the competitive process but also to guarantee the conditions for innovation and growth of new services.

References

- Arvech H. and L. Johnson, 1962; "Behavior of the Firm Under Regulatory Constraint", *American Economic Review*, Vol. 52, 1052-1069.
- Baron D. and R. Myerson, 1982. "Regulating a Monopolist with Unknown Costs," *Econometrica*, vol. 50, 911-30.
- Boiteux, M. 1956. "Sur la gestion des monopoles publics astreints à l'équilibre budgétaire", *Econometrica*, 24, 22-40.
- Brock G.W., 2002. "Historical Overview", in *Handbook of Telecommunications Economics*, édité par M. Cave *et al.*, North Holland.
- Chaaban, J., 2008. "Using Cost Allocation to Partially Regulate Multi-market Utilities", *Competition and Regulation in Network Industries*, Vol. 9.
- Dana J. D. and K.E. Spier, 1994. "Designing a private industry : Government auctions with endogenous market structure", *Journal of Public Economics*, 53, 127-147.
- Faure-Grimaud A., 2002. "Using Stock Price Information to Regulate Firms", *Review of Economic Studies*, vol. 69, 169-190.
- Faure-Grimaud A. and D. Martimort, 2003. "Regulatory Inertia", *RAND journal of Economics*, Vol. 34, 413-437.
- Fumagalli C. and M. Motta, 2010, "A Simple Theory of Predation", mimeo.
- Hausman J., Pakes A. and G. Rosston, 1997: "Valuing the Effect of Regulation on New Services in Telecommunications" *Brookings Papers on Economic Activity. Microeconomics*.
- Jullien B. and P. Rey, 2006, "Charges de terminaison et concurrence: quelques leçons de la littérature économique", Rapport IDEI #6
- Jullien B. and P. Rey, 2008, "Notes on the economics of termination charges", Rapport IDEI #8.
- Kornai J., 1986. "The Soft Budget Constraint," *Kyklos*, Vol. 39, 3-30.

Laffont J.J., 1996. "Industrial Policy and Politics," *International Journal of Industrial Organization*, vol. 14, 1-27.

Laffont J.J., P. Rey and J. Tirole, 1998a. "Network Competition: I. Overview and Nondiscriminatory Pricing," *RAND Journal of Economics*, vol. 29, 1-37.

Laffont J.J., P. Rey and J. Tirole, 1998b. "Network Competition: II. Price Discrimination," *RAND Journal of Economics*, vol. 29, 38-56.

Laffont J.J. and J. Tirole, 1986. "Using Cost Observation to Regulate Firms," *Journal of Political Economy*, Vol. 94, 614-41.

Laffont J.J. and J. Tirole, 1991. "The politics of government decision making: A theory of regulatory capture", *Quarterly Journal of Economics*, Vol. 106, 1089-1127.

Laffont J.J. and J. Tirole, 1993. *The Theory of Incentives in Regulation and Procurement*. Cambridge, Mass.: MIT Press.

Laffont J.J. and J. Tirole, 2001. *Competition in Telecommunications*, collection Munich Lectures, Cambridge, Mass.: MIT Press.

Salant D., 1995 "Behind the Revolving Door: a New View of Public Utility Regulation", *RAND Journal of Economics*, Vol. 26, 362-377.

Schmidt K., 1996. "The Costs and Benefits of Privatization: An Incomplete Contracts Approach", *Journal of Law, Economics and Organization*, vol. 12, 1-24.

Segal I., and M. Whinston, 2007. "Antitrust in Innovative Industries" *American Economic Review*, Vol. 97: 1703–1730.