Abstract

The concept of natural monopoly presents a challenging public policy dilemma. On the one hand a natural monopoly implies that efficiency in production would be better served if a single firm supplies the entire market. On the other hand, in the absence of any competition the monopoly holder will be tempted to exploit his natural monopoly power in order to maximize its profits.

This chapter will take a closer look at the model of natural monopoly. It will address those areas where an unregulated natural monopoly is generally considered to be the cause of concern, before offering a brief overview of the regulatory process and some of its specific regulatory tools. It should be noted that this chapter mainly aims to provide an introduction to the vast literature on this topic, which has fascinated many economic and legal scholars over the years.

JEL classification: K23, L90

Keywords: Cream Skimming, Price Regulation, Incentive Regulation, Contestable Markets, Public Ownership

1. The Natural Monopoly Model

A natural monopoly exists in an industry where a single firm can produce output such as to supply the market at a lower per unit-cost than can two or more firms.

The telephone industry, electricity and water supply are often cited as examples of natural monopolies. These industries face relatively high fixed cost structures. The costs necessary to produce even a small amount are high. In turn, once the initial investment has been made, the average costs decline with every unit produced. Competition in these industries is deemed socially undesirable because the existence of a large number of firms would result in needless duplication of capital equipment. The classic example might be that of two separate companies providing local water supplies, each constructing underground pipelines.
In undergraduate textbooks one finds the natural monopoly condition linked to the issue of economies of scale. Traditionally, natural monopoly is often described as a situation where one firm may realize such economies of scale that it can produce the market’s desired output at an average cost which is lower than two firms could with smaller scale processes.

The modern approach to defining natural monopoly was initiated by William J. Baumol (1977). In an industry where a single firm can produce output to supply the entire market at a lower per-unit cost than can two or more firms (subadditivity of the cost functions) or in an industry to which entrants are not ‘naturally’ attracted and are incapable of survival, even in the absence of predatory measures by the incumbent monopolist (sustainability of monopoly) the single firm is called a natural monopoly.

The concept of subadditivity is a precise mathematical representation of the natural monopoly concept (Baumol, 1977). If all potential active firms in the industry have access to the same technology, represented by a cost function $c$, then at an aggregate output $x$, the industry is a natural monopoly if

$$ c (x) \leq c (x^1) + \ldots + $$

for any set of outputs of $x^1, \ldots, x^t$ such that

$$ \sum_{i=1}^{t} x^i = x $$

A cost function $c$ is globally subadditive if for any non-negative output vectors $x$ and $y$, $y$,

$$ c (x^1 + y^1, \ldots, x^n + y^n) \leq c (x^1, \ldots, x^n) + c (y^1, \ldots, y^n) $$

Given the production of a set $N \setminus \Lambda, ..., n\Lambda$ of indivisible objects, the cost function is subadditive if

$$ c (S \cup T) \leq c (S) + c (T) $$

for any disjoint subsets $S$ and $R$ (on the concept of subadditivity, see Baumol, 1977; Sharkey, 1982).

There is a close relation between economies of scale and subadditivity. In a single-product firm economies of scale are a sufficient but not necessary condition for a natural monopoly. Production processes, although subadditive at output levels which exhibit economies of scale (decreasing average costs), may also be subadditive when exhibiting increasing average costs. In a multi-product firm, for instance, economies of scope may create a natural monopoly although there are no economies of scale with regard to the production of the goods separately.
Herein lies the difference between a strong and a weak natural monopoly (Gegax and Nowotny, 1993, p. 67). While strong natural monopolies exhibit decreasing average costs, the weak natural monopoly firm exhibits increasing average costs even though its costs are subadditive. The latter finds itself in a situation where it may not be able to prevent entry by other firms, for example when its monopoly position is non-sustainable. Since costs are subadditive, society might prefer a market consisting of only one firm rather than a multiplicity of firms producing the same output. Therefore, in the case of a weak natural monopoly it is often considered desirable that regulatory authorities bar entry into the market of the weak natural monopoly and in turn regulate prices.

Graphically, a strong natural monopoly exists when the long-run average cost curve of a single firm is still declining at the point where it intersects the total market demand curve for the product. The D line on Figure 1 represents the demand curve, which is also the market-demand curve. A monopolist will supply the market with output at a price $P_o$. The optimal scale of the firm is reached at point A in the figure.

The market conditions, illustrated in Figure 1, allow for one firm to provide the entire market cheaper than could two or more firms. Suppose the market consists of four firms, each producing at an optimal level, where marginal revenue equals marginal costs. Thus, each firm produces a quantity of 1000. The total market demand curve determines price, which adds up to $P_o$. At scale $B$, where the market is restructured into 4 firms, instead of one monopoly firm, each firm produces at a smaller scale. The total amount produced is unchanged but the average unit costs are higher at scale $B$. The market allows for one producer to generate total market demand at cheaper costs than can two or more equal producers.

Figure 1
A. Why Regulate Natural Monopoly?

Many supposed natural monopolies are the subject of various types of regulation. As described above, under conditions of natural monopoly the market is best served when one firm supplies total market demand. Public interest theory claims to provide an explanation for government intervention in what may be considered a market imperfection. The need to avoid duplication of facilities, particularly fixed costs, would serve as a justification for traditional entry regulation. Consider in this respect the restructuring of the telecom industry in the United States that broke up the Bell system to AT&T, while forbidding Regional Bell Operating Systems to enter the lines of business assigned to AT&T in order to prevent destructive competition (Crandall, 1988).

In Part A we will try to provide a brief insight into some other motives behind the regulation of natural monopolies. We refer to Chapter 5940 in this volume for a detailed review of the specifics of regulation of utilities and to Chapter 5000 for a general discussion of the various theories of regulation.

2. Allocative Inefficiency

Under perfect competition prices of goods equal marginal cost, as firms engage in a competitive bidding process. Under conditions of monopoly, the profit-maximizing behavior of the incumbent firm will lead to a higher price charged to consumers and a lower output. It enables the seller to capture much of the value that would otherwise be attained by consumers. Monopoly pricing thus results in a wealth transfer from consumers of a product to the seller. At the higher price, at which the monopolist tries to maximize profits, a group of potential consumers will be excluded as they will not be able to afford the product at the higher (artificially set) price. Thus, monopoly leads to the classic case of the occurrence of dead weight losses: the part of the consumer surplus that the monopolist cannot appropriate but consumers lose.

Now as a result of the monopoly pricing scheme, these consumers may be forced to consume more costly substitutes or less useful products, although society’s resources would be better used producing more of the good provided by the monopoly firm. Furthermore, the argument goes that by limiting output the monopolist underutilizes productive resources.

The argument of the negative consequences of monopoly on economic welfare has been the subject of heavy debate. This article will not venture into the broad discussion of welfare economics, monopoly and distributive justice (for an introduction, see Tullock, 1967; Rahl, 1967; for case studies on the consequences of monopoly pricing and welfare, Albon, 1988). We can,
however, focus on a few of the arguments concerning the case of *natural* monopoly which challenge the relevance of the alleged allocative inefficiency. The classic opposition to monopoly rents as opposed to everyday rent-seeking by the common man is that monopoly rents are the result of an artificial scarcity rather than a natural scarcity (Schap, 1985).

The question arises whether the same really can be said about an unregulated natural monopolist. Early on, Posner (1969) rightly noted that market power in the latter case stems from cost and demand characteristics of the market, not from unfair or restrictive practices.

### 3. X-Inefficiency

The condition of natural monopoly raises the question whether internal efficiency, cost minimization by the firm, is achieved under natural monopoly. Does a monopoly firm put its resources to the best possible use within the existing state of technology?

Modern antitrust economists have used the term ‘X-Inefficiency’ to indicate the internal wastes that occur when a firm acquires monopoly power and is no longer pressured by strong competitors to keep its costs at the competitive minimum. Often-cited legendary examples are US Steel, General Motors, Sears, IBM and American Airlines. These giant firms, which once dominated their industries, are accused of falling victim to their own inefficiencies (Mueller, 1996). Empirical data suggest that the amount to be gained by increasing X-efficiency is significant (for a review, see Leibenstein, 1966).

Generally, in a competitive market firms have an incentive to reduce costs, in order to obtain higher profits by selling at the same price or a price between the old price and the new cost level. Although cost reduction might be short-lived in a competitive situation, as competitors reduce their production costs and adjust their prices to those of their direct competitors, the concern for survival provides a firm in such a market with a strong incentive to minimize costs (Dewey, 1959). If a firm fails to anticipate or match the cost reductions of its competitors, it might suddenly find itself in a market dominated by its competitors. Where there are no significant entry barriers the threat of potential competition will hold price down to cost. Otherwise other firms will enter the market at the same scale of production, sell at a slightly lower price and capture the whole market for as long as it is profitable (for more on contestability, see Section 8).

Also, it could be argued a monopoly firm has an even stronger incentive to minimize costs in order to gain maximized profits. Since the threat of a counter
reaction to its pricing schemes is absent, it does not face the risk that the consequential benefits will only be short-lived.

4. Technological Progress

Technological developments have been the drive behind the transformation of certain natural monopoly markets to more competitive outcomes. Most notably, this is the case for the more recent changes in the telecommunications industry, where the enhancement and development of microwave and satellite technology has come to provide a strong substitute for the traditional cable networks. The value of technical development should not be underestimated. Technological progress often reduces production costs or creates new products and has been of enormous importance to economic welfare.

The classic argument goes that monopoly firms lack an efficient incentive to promote technical change and invest in expensive R&D programs. Allegedly, a monopoly firm would discourage progress. By virtue of its protected position it would not fear that a rival will promote products and production methods and would therefore not be driven to pioneering himself. Real-life observations regarding the introduction of new technology in monopoly firms seem to validate this criticism - witness the long life of equipment in telephone industries. This is often the case, regardless of whether the monopoly firm is conducted as a public or private monopoly (Dewey, 1959). Some empirical data suggests that small, profit-seeking firms are responsible for most major innovations (Scherer, 1984).

However, there are strong arguments that provide indications that contradict the traditional allegations concerning the case of under-innovation under monopoly.

Even when assuming that a monopoly firm will not introduce new products unless the cost of the new product is less than the marginal cost of the old (as sunk costs are bygones), there is no reason the same could not be said about competitive firms (Fellner, 1951).

Furthermore, an important point has to be made. The fact that an industry is a monopoly does not mean that only one firm is pursuing research and development in its technology. Through the presence of external forces it is likely that the incumbent monopoly firm will feel the pressure to spend time and money on innovation, in order to safeguard its position (Posner, 1969). Certainly in an unregulated market, with free entry, successful research in the field of production methods could seriously threaten the position of natural monopoly firms. Although a natural monopolist is less concerned about survival, the possible threat of the introduction of new technology in substitute
markets should provide monopoly firms with a strong incentive to anticipate such developments through R&D expenditure (this relates to the theory of contestable markets, see Section 8).

At the other end of the spectrum Schumpeter (1950) holds that there is a positive relation between innovation and market power. A monopoly firm would be a strong instead of a poor innovator. Superior access to capital, the ability to pool risks and economies of scale in the maintenance of R&D laboratories are likely to advance industrial technology.

5. Cream-Skinning and Cross-Subsidization

Because of social considerations, government may often feel the need to ensure the provision of certain products or services at a lower-than-cost price to some consumers. For instance, it is quite common that governments demand the provision of ‘universal services’ to consumers by telephone companies, the availability of minimum services at reasonable prices, even to small and distant communities where the small scale of operation may lead to very high costs, which often results in the occurrence of losses.

The delivery of such services is often financed by obtaining higher profits on the sale of other products and services. This is termed cross-subsidization, referring to the practice where the difference between the price charged to the targeted consumers and the cost of supply might be funded by cross-subsidization from the prices paid by other consumers or, in a multiproduct firm by the purchases made for other products or services.

When a firm, burdened with ‘universal service’ obligations of some sort, is not protected from price competition and free entry of competitors, it might not be able to maintain its position in the market as potential competitors will enter these market segments that are provided at prices well above cost. Cream-skimming occurs when a supplier concentrates only on those areas of the market where the costs of supply are lowest, for instance because of geographical reasons (on cream-skimming in a natural monopoly market, see Zupan, 1990). Regulated firms on such a market, with universal services obligations, might find themselves in a possibly fatally detrimental situation, when competitors capture these low-cost/high-profit parts of the market which are crucial to recover losses made in the high cost market parts. Take the example of the US Postal Service. Without legislative protection, its uniform pricing scheme, according to which it charges the same price to deliver a letter anywhere in the United States regardless of distance and specific difficulties, would soon deprive it of its most profitable routes.
Yet, the practice of cross-subsidization, when a company that supplies more than one product and uses the revenues from product A to recover a portion of the costs of product B, generates economic inefficiencies. Although benefits from economics of joint production might have to be sacrificed, in most cases the consumers of product A would be better off if the products were produced and priced separately, while the customers of good B are given incorrect price signals about the incremental costs of producing product B (Spulber, 1995). The cross-subsidization of loss-making services by taxing remunerative services has been demonstrated to lead to significant allocative inefficiencies (Brennan, 1991).

B. The Regulatory Process

6. Price Regulation

Price control, although driven to the background in the years of deregulation, has been of increased importance in the recent trend of privatization in Europe. From a public interest perspective, price control should allow regulators to set prices at a level which induces allocative and productive efficiency. This part provides a brief, non-technical introduction to some of the tools governments have at their disposal to assure that firms meet consumer demand at efficient cost levels. For a more in-depth look at the different forms of price regulation and its analysis, reference is made to Chapter 5200, Price Regulation, which also deals with natural monopoly.

Figure 2
Marginal and Average Cost Pricing and Rate of Return Policy

Without regulation, $Q_m$ in Figure 2 represents the profit-maximizing output of the monopolist and the demand, in turn, determines the market price $P_m$. The monopoly earns a positive economic profit represented by the area of the rectangle $P_mFGH$. Welfare maximization to society as a whole is achieved at the quantity-price of $P_1$, $Q_1$. In other words, regulating authorities should set a price $P_1$ (marginal cost pricing) in order to maximize economic welfare. However, at price $P_1$, consumers will buy a quantity of $Q_1$, whereas $AC$ is $NQ_1$, which is greater than the price $P_3$. This will result in a total negative economic profit, shown by the area of the rectangle $P_1P_3AM$. In the long run, the monopoly firm will not stay in business. If the commodity or service provided is desirable, the only way to keep the monopoly firm in business is to provide a public subsidy to the amount of $P_1P_3AM$.

The political problems associated with subsidization, its implementation and financing and the difficulties of calculating demand and $MC$ have led to the application in the public utility field of average cost pricing. In Figure 2 such a price is $P_2$, determined by the intersection of the demand and long-run $AC$ curve. The output under average-cost pricing, $Q_2$, is greater than the unregulated monopoly output of $Q_m$. Also, part of the welfare costs arising from restricted output by an unregulated monopoly is eliminated. Expansion of output, from $Q_m$ to $Q_2$ provides benefits to consumers that are greater than the additional costs.

On the other hand, average cost pricing can hardly be deemed entirely satisfactory either. Under average cost pricing, when $Q_2$ is produced, welfare losses are caused because at this point average costs (the $AC$ curve) exceed marginal costs ($MC$). Graphically, the $AMO$ triangle in Figure 2 represents this consequential welfare loss. When applying a rate-of-return policy, regulating agencies focus on the rate of return on invested capital (accounting profit) earned by a monopoly (fair rate of return) (Moorehouse, 1995). Allowing regulated firms to acquire a total sum that consist of annual expenditure plus a reasonable profit on capital investment, the so-called ‘fair’ rate of return, was constructed by American courts and the regulating bodies in order to meet constitutional demands of utilities to set prices on a ‘just and reasonable’ level.

This can be formulated as $E + (r \cdot RB)$ where $E$ represents the firms annual expenditure, $r$ is the multiplier, representing the fair rate of return, and $RB$ the rate (attributed value of the capital investment).

If the realized rate of return is higher than what is considered to be a normal return, then the price must be above average cost. In a trial-and-error fashion regulators try to locate the price where profit is normal, for example, where price equals average cost.
Allowing the regulated monopolist a fair rate of return creates various economic problems that have to be taken into account. Auditing costs involved in determining the firm’s capital base are considerable. Especially the determination of $r$, which should reflect a level of return that is satisfactory to attract investment, is problematic. Looking at other ‘comparable’ industries or applying the capital asset pricing model, where one looks at the returns obtained by investors from a portfolio of investments, as modified by the difference between the returns from shares in utilities and those from more general market shares, it is clear that these are imperfect methods for determining a rate of return that potential investors will demand from the regulated industry.

Also, the perverse effects on incentives that occur when applying the standard rate of return policy are perhaps even more troublesome (Train, 1991). The regulated firm has an incentive to inflate its capital cost figures, since higher costs imply higher absolute returns (Baron and Taggart, 1977, on profit-maximizing behavior under rate of return regulation, see Hughes, 1990). If the regulator sets the fair rate of return above the cost of capital, the regulated firm is likely to utilize more capital than if it were unregulated. Thus, it might use inefficient high capital/labor ratios for its output. Also, average cost pricing diminishes the monopolist’s incentive to minimize costs. Figure 2 illustrates some of the consequences of this behavior. Inflation of costs shifts the actual \( AC \) curve to \( AC' \). At a price of \( P_2 \), losses occur, therefore, regulators grant a price increase to cover the higher costs (point \( C \)).

When implementing rate-of-return regulation, the complexity accompanied with instituting and enforcing regulation schemes may often lead to delays in the regulatory response to changes in costs and other market conditions. When costs are falling during a certain period of time firms will benefit from these so-called regulatory lags, as they will be able to enjoy rates of return greater than those deemed ‘fair’ in the outcome of the regulatory proceedings. Several authors have argued in favor of regulatory lags: they would, at least temporarily, provide a way to allow regulated firms to profit from achieving lower costs (Baumol, 1967; Williamson, 1971; Bailey, 1973). However, if costs shift upwards, for instance due to factors external to the regulated firm, the opposite effect will occur and, the firm will incur costs that were not foreseen by the regulator.

Also, because profit rates are restricted, the firm’s managers face the option of providing themselves with amenities in their salaries or the firm’s profits. Whether they do take advantage of this option in practice is largely dependent on whether they are externally controlled by reliable public officials, which in turn involves considerable monitoring costs.
Price Discrimination
Charging consumers different prices relative to what they are willing to pay, even though the costs of producing and supplying the goods or services are the same, has been demonstrated to enable allocative efficiency. With this technique, often referred to as Ramsey pricing, information on the price-elasticity of the different goods should allow for efficient price setting (the higher the price-elasticity, the closer the price needs to be set to marginal cost) (Ramsey, 1927). The objection often heard is that such a pricing scheme involves a wealth transfer from the consumer (consumer surplus) to the producer. Also, severe price discrimination, often termed predatory discrimination, is an effective method by which competition may be crushed out (see however for a detailed and comprehensive account McGee’s case study of the Oil of Indiana case, 1958).

The classical analysis of price discrimination was set out by Pigou (1920). A distinction between three degrees of discrimination was made. The first degree involves a different price set for every unit purchased by every consumer, in such a way that virtually all the possible consumer surplus is obtained by the producer. Although this pricing mechanism can be efficient, as marginal decisions are made as marginal cost, it is often opposed on income distributional grounds: the transfer from consumer to producer. Second-degree price discrimination consists of pricing groups according to their willingness to pay, where all those with a demand price above a certain level are charged one price, while those with a lower demand price are charged a lower price. Third-degree price discrimination comes into effect when consumers are divided into separate groups, each group is charged a different monopoly price. This technique is of course strongly dependent on the possibility for the seller to identify groups in each specific case, which will vary according to market circumstances.

In the telephone and electricity industries, for instance, the distinction is made between residential and commercial customers, who pay different prices (Hollas and Friedland, 1980; Primeaux and Nelson, 1980; Naughton, 1986; Eckel, 1987). Sometimes, sorting consumers by their willingness to pay can be achieved by requiring customers to tie (Cummings and Ruhter, 1979) or bundle (Adams and Yellen, 1976). In circumstances of uncertainty of demand, where prices must be established before demand is known, special distinctions, such as ‘saver-airfares’ requiring early bookings may be enforced on separate consumers (Leland and Meyer, 1976; Sherman and Visscher, 1982).

Peak-Load Pricing
When demand follows a periodic cycle, during which demand might be high at certain times and low at others, peak-load pricing might offer a way to achieve marginal cost pricing. As marginal cost generally rises with output,
price variations will allow it to reflect the higher costs. This allows for the moderation of the demand cycle while establishing a more effective use of capacity (Crew, Fernando and Kleindorfer, 1995). Higher pricing during periods of peak demand might discourage use and save costly capacity, whilst lower prices when demand is low might encourage use of capacity that otherwise would have been left idle.

7. Incentive Regulation

Many of the problems arising from the pricing models, such as the Averech-Johnson model find their origin in the absence of incentives to operate at minimum cost levels. The motivation behind incentive regulation is to provide the firm with the motivation to behave more consistently with regard to the social optimum.

Price-Cap Regulation or RPI-X

Requiring the firm to increase its prices for each year within a given period by no more than the retail price index (RPI) minus a variable factor ($X$) which is the agency’s assessment of the firm’s cost-efficiency potential, price-cap regulation was found to be superior to the fair rate of return method both in terms of efficiency and administrative costs (Littlechild and Beesley, 1992). The system, as described above, would require less information from the firms to the regulating bodies, which would not only make this model less costly but also diminish the capture problems associated with rate of return regulation (for an overview, see Chapter 5200, Price Regulation).

C. Alternatives for Regulation

The inefficiencies (Coase, 1959, Posner, 1969, 1975) and costs (Gerwig, 1962; Hahn and Hird, 1991) of the regulatory processes have been well documented over the years. Many scholars became dissatisfied with the public interest approach to regulation and looked for an alternative that would explain data that did not correspond with the normative approach to regulation. Capture theory, which holds that producers are the winners under most, began to shape (Jordan, 1972). A considerable body of literature, the *economic theory of regulation*, grew from this (Peltzman, 1976; Stigler, 1971, 1976; for an overview, see Viscusi, Vernon and Harrington, 1995).
This in turn, has given rise to new, appealing theories where the approach is to have optimality induced without regulation, even in markets with one producer. Competition amongst numerous firms can be used to achieve optimality under natural monopoly conditions, where competition is held amongst firms that could produce. The argument goes that the pressure from potential producers will induce efficient pricing decisions by the natural monopolist.

8. Contestable Markets

Building upon Demsetz’s ideas (1968) on the threat of potential competition as a disciplinary mechanism, the theory of contestable markets was further developed by Willig (1980) and Baumol, Panzar and Willig (1982) and applied to natural monopoly (Coursey, Issac and Smith, 1984). In a market where entry and exit are completely free and unconstrained, the threat of potential competition may hold price down to cost. When potential entrants exercise strong constraints on the behavior of the incumbent monopoly firm, the latter will be moved to pricing schemes closer to cost. Firms in a contestable market will not be able to gain pricing profits that exceed normal profits under competitive circumstances, as otherwise other firms will enter the market at the same scale of production, sell at a slightly lower price and capture the whole market for as long as it may be profitable, a practice often referred to as hit-and-run entry and exit (Baumol, 1982). When entry is allowed, new firms will be able to enter the market of a Ramsey-pricing natural monopolist at a lower price (Faulhaber, 1975; Sharkey, 1982). Thus, when the nature of the market allows for only one firm to provide total demand, competition can assert itself by deciding which firm will obtain market domination.

Contestability of a market would render regulation pure waste, as without regulation the monopoly would yield price efficiency. Instead, given this theory, when determining the proper form of regulation, governments should restrain from measures that obstruct potential entry and create an environment that promotes contestability. In other words, regulators should encourage rather than prevent entry into the natural monopoly market.

However, in order for a market to be qualified as contestable various conditions, most notably the absence of any significant entry barriers, are demanded. Perfectly contestable markets are based on free entry - the firm does not have to incur any cost that is not also incurred by the incumbent natural monopolist - by entrepreneurs who do not face any disadvantages in relation to the incumbent monopolist. It assumes the possibility for the firms to freely exit, that is, potential recoupal of all costs incurred upon entering - minus
depreciation. Also, the potential entrants must have access to equally efficient production technology as employed by the incumbent firm. Finally, the incumbent firm must not be able to adjust prices instantaneously when faced with the threat of entry.

As a theoretical construction the theory of contestable market certainly has its merits. Indeed, a natural monopoly firm may not be immune to profit-seeking hit and run entry (Faulhaber, 1975; Baumol, Bailey and Willig, 1977; Baumol, Panzar and Willig, 1982; Sharkey, 1982). However, it remains uncertain whether perfectly contestable markets do actually exist. The assumption which holds that an entrant can leave a market without costs when its presence is no longer profitable is rarely encountered in real-life economics (Waterson, 1988). Also, sunk costs, the difference between the \( \text{ex ante} \) opportunity cost of the funds and the value that could be recovered \( \text{ex ante} \), have early on been demonstrated to deter entry (Eaton and Lipsey, 1978; Spence, 1977, 1980; Dixit, 1980). Cost curves reflecting large sunk fixed costs, already borne by the incumbent firm, place potential entrants in a disadvantaged situation. (Bailey and Panzar, 1981).

9. Entry Regulation and Auction

Until 1968, the inevitability of regulation of natural monopoly was broadly accepted. It was Demsetz (1968) who argued that formal regulation, as described above, is uncalled-for where government can allow ‘rivalrous competitors’ to bid for the exclusive right to supply a good or service over the given ‘franchise period’. Monopoly pricing is prevented, as competition has asserted itself at the bidding stage. In other words, competition at the franchise stage will be sufficient to reduce the price below that of the monopoly profit maximizer. Therefore, a monopoly structure does not necessarily lead to monopoly behavior. Along the same lines we may situate Chadwick’s (see Crain and Eklund, 1976) historical reasoning in his investigation of water supply in London in the 1850s, where he advocates competition for the market rather than the costly and that time prevalent competition within the market. Successful and competitive bidding, in terms of prices and services offered, could initially eliminate the need for a principal-agent relationship (Chadwick, 1859).

In certain cases, government might find it appropriate, most notably in the case of a natural monopoly, to limit the number of firms operating in a certain market.

Bidding refers to a free and open right to supply the market, where the regulator announces that it will accept bids from all firms that are willing to provide the goods. Each bid could consist of the price that the firm would agree to charge consumers when awarded the franchise. In such a bidding process,
price will eventually be bid down to a price at which the winning firm earns zero-profits. This will be the result of a repeated process where firms given the choice between zero profits from not winning the auction, and earning a small but positive profit by bidding below the lowest bid, will choose the latter. Even in the case where the bidding would be held at once, each firm would realize it could only win if it bids at a price that provides zero economic profit with least-cost production. (Coursey, Isaac and Smith, 1984; for a survey on the bidding process in the electricity industry, see Rozek, 1989; in cable television franchise allocation, see Zupan, 1989). This form of bidding, often referred to as Chadwick bidding, should be distinguished from auctions where the state is acting as a monopolist attempting to attain high prices, for instance when auctioning oil exploitation rights, instead of acting as a consumer agent (Posner, 1972; Waterson, 1988).

At this point, reference should be made to recent proposals to have the auctions centered around an incentive contract, where the winner is consequentially bound to a desirable incentive arrangement (McAffee and McMillan, 1987; also Riordan and Sappington, 1987).

However, the effective pursuit of economic welfare through the competitive awarding of monopoly franchises has its difficulties. Auction theory suggest that the quality of the winning bid only increases as the number of bidders increases (for a different view, see Bishop and Bishop, 1996). As has been demonstrated early on (see Williamson, 1976) in the cable television industry, the winner of a monopoly franchising procedure does not always fulfill the service contract as promised. Here, monitoring costs and the principle-agent problems should be taken into account. Government contracts are often won by bidding in terms that cannot be met (Sherman, 1989), the main reason being that the services are complicated and all eventualities cannot be anticipated fully when contracts are drawn up (see also the debate on the winner’s curse, Thaler, 1992). Also, once a firm has obtained the franchise for a certain period of time with exclusive information on costs, production processes and control over resources, future franchise biddings will be among unequals. Furthermore, it should be noted that costs and demand change over time in such a way that the price fixed in the franchise contract might not be optimal at a later point of time (Williamson, 1976). Contingency clauses, anticipating future events might offer a remedy, but are merely second best solutions, as they are bound to bring about considerable imperfections. Contractual stipulations, which specify the price movement in the event of price changes, demand information from a government which it simply does not have. Another type of contingency clause, the implementation of procedures by which to revise prices periodically, faces the same problem of asymmetry of information, as it provides the regulated firm with incentives to report smaller than real cost changes, for instance from
technological progress (Train, 1991, p. 301). Reference can be made to the suggestion (Posner, 1992) to have repeated auctions held with short-term contracts for the winning firm. These short-time contracts will allow a better reflection of changes in cost and demand. However, it should be reminded that these auctions will only result in efficient price offerings if the incumbent firm has no other advantages with respect to the other firms involved in the franchise bid.

Loeb and Magat (1979) propose an institutional arrangement consisting of a mixing of regulation and franchising, while supposedly eliminating the problems associated with both. As an alternative to rate-of-return regulation, they set forth a decentralized system of regulation in which the utility chooses its own price and where the regulatory agency subsidizes the utility on a per unit basis equal to the consumer surplus at the selected price. The classic opposition to subsidies, that revenues to provide a subsidy have to be collected elsewhere and that any subsidy policy as such brings along allocative inefficiencies, are countered by the introduction of a sale of the franchise which should reduce the net subsidy or, alternatively, the imposition on the utility of a lump-sum tax by the regulatory agency in order to recover part of the subsidy. Sharkey (1979), commenting on the Loeb-Magat (L-M) scheme, argues that such a policy might hold if conditions exist that the net subsidy paid to the utility is sufficiently small. Aside from imperfections of the regulatory bodies and the costs resulting from political manipulation, the L-M scheme is less convincing where one attaches more importance to the quality of service as opposed to the price. Nevertheless, as an alternative to rate of return regulation the L-M model has the advantage that the regulatory body does not need to obtain or verify information about the cost function of the utility and as such it has considerable merit (Sharkey, 1979). Moreover, experimental research on the behavioral robustness of the contestable market hypothesis has shown it to be promising (Coursey, et al., 1984; Harrison and McKee, 1985) especially in environments where the Bertrand-Nash assumption is satisfied (Harrison, 1987).

10. Public Ownership

Another possible alternative to regulation is government ownership of enterprises that provide services under conditions of natural monopoly. Although one could argue that regulation occurs in its most severe and complete form under public ownership, its unique characteristics distinguish it from the broad range of traditional regulatory processes (Ogus, 1994). Instead of having a privately owned monopoly with profit-seeking shareholders one could institute a publicly owned enterprise with less concern about profits. This
lack of profit-maximizing incentives in a public enterprise is sometimes thought to be beneficial, as it allows publicly responsible attention to non-financial goals and/or distributional goals (on the various possible reasons for state ownership, Ahoroni, 1986). The institutional framework of public ownership would provide a way to impose public interest prices and standards. This would allow the equation of prices to marginal cost, to have monopoly profits avoided, and so on.

Half a century ago, government ownership of firms was thought of as the key solution to market imperfections, such as monopoly. Due to the failures of competition, the regulation during the Great Depression, the apparent success of Soviet industrialization and a misunderstanding of the consequences of political control over firms, the state took control of significant parts of production in the economy of many countries.

In the last 29 years, a renewed faith in the market process induced governments in market economies throughout the world to privatize most of their sectors, including strategic ones such as steel, energy and telecommunications.

The absence of a profit incentive under the institutional framework of public ownership had proven to be a high price to pay. In a public enterprise which lacks a group of residual profit-claiming shareholders, who emphasize fiscal goals and enforce efficient performance through management, economic efficiency is no longer guaranteed (Spann, 1977; Williamson, 1987; for a survey of economic performance under public ownership in the British fuel and power industry see Shepard, 1965). When assets are publicly owned, the public manager has relatively weak incentives to reduce costs or to improve quality or innovate, because he only gets a fraction of the return as a non-owner (Hart, Shleifer and Vishny, 1997; more on public managers, De Alessi, 1974, 1980). In private corporations, the shareholders’ ability to sell their vote or to vote out management creates incentives for management to serve the interests of the owner. The diffuse, non-transferable shareholding that characterizes government ownership reduces these incentives. Those in control of the enterprise pay less attention to the taxpaying shareholders and are more likely to succumb to more concentrated interest groups, such as suppliers, consumers, employees, and so on (Zeckhauser and Horer, 1989).

The statutory monopoly will become the primary source of information about industry possibilities. The monopoly will not suffer as a competitive firm would when it is wrong, because regulators either cannot appreciate its errors completely or will forgive them. Regulatory agencies, distanced from the industry, might have a hard time to reflect the complexity of the industry. As regulators cannot evaluate all decisions, inefficient technologies may be chosen for years (Sherman, 1989).
Also, the question remains what goals replace the profit incentive? Imposition of vague goals often results in diminished accountability which imposes the risk of inefficient results. Moreover, it should be noted that the absence of a profit incentive does not guarantee the absence of monopoly prices; these remain a possibility, for instance to make the life of managers easier, and so on.

A public enterprise is often constrained by its budget. As it is required only to have a certain difference between total revenue and total costs, it may, especially when no stronger restrictions have been set out, seek to maximize its total expenditures or its revenue. In order to do this it may well follow a monopoly pricing rule. The US Postal Service can serve as an example of a public enterprise (since its 1970 charter) exhibiting such behavior, empirical evidence shows it has behaved as a cost or revenue maximizer for many years (Sherman, 1989, pp. 265-268). A publicly-owned firm with limited restrictions such as budget constraint has considerable managerial discretion, which often leads to the pursuit of various goals such as budget maximization (Niskanen, 1971; Crew and Kleindorfer, 1979), revenue maximization (Baumol, 1976) and the maximization of total output.

The importance of innovation in market economies should not be disregarded. Voices as early as Marshall (1907), have argued that government is generally a poor innovator. The development and adoption of new technologies in telecommunications, that occurred shortly after the privatization of phone companies in the US, may serve as an example of the benefits of private ownership (Winston, 1998). In industries where innovation is crucial, the case for government provision is strained (Shleifer, 1998). Technology changes often have enormous impact on cost structures and may cause a natural monopoly market to move to a more competitive setting.

Selected Bibliography on Regulation of Natural Monopoly (5400)


Regulation of Natural Monopoly


Brown, Lorenzo (1988), ‘Modelling Diversified Utilities and Cross Subsidization’, 10 Resources and


Chadwick, Sir Edwin (1859), ‘Results of Different Principles of Legislation and Administration in Europe: Competition for the Field, as Compared with Competition Within the Field of Service’, 22 Journal of the Royal Statistical Society, 384-420.


Regulation of Natural Monopoly

Regulation of Natural Monopoly

University of Miami Law Review, 433 ff.
Mechanisms', 34 Kyklos, 388-404.


Georgantzis, Nikolaos (1992), The Social Aspects of Telecommunications: Targeted Subsidies as a Means to Achieve Universal Basic Services, Report for the A-3 Unit (Economic Questions and Studies) of DG IV (Competition) of the European Commission.


Goldberg, Victor P. (1979), 'Protecting the Right to be Served by Public Utilities', 1 Research in Law and Economics, 145-156.


Kelly, Kevin et al. (1987), Some Economic Principles for Pricing Wheeled Power, Columbus, National Regulatory Research Institute, 361 p.


Mann, Patrick C. (1979), ‘The Dynamics of Traditional Rate Regulation’, 1 Research in Law and Economics, 195-212.


McNicol, David L. (1979), 'Price Discrimination and Peak-load Pricing Subject to Rate of Return Constraint', 1 Research in Law and Economics, 213-238.


Niskanen, W.A., Jr (1971), Bureaucracy and Representative Government, Chicago, Aldine
Atherthon.
Poole, Robert W., Jr (ed) (1985), Unnatural Monopolies: The Case for Deregulating Public
Regulation of Natural Monopoly


Samuels, Warren G. and Mercuro, Nicholas (1976), ‘Property Rights, Equity, and Public Utility
Public Utility Papers, Institute of Public Utilities, Michigan State University, 44-82.
Sappington, David E.M. (1980), ‘Strategic Firm Behavior under a Dynamic Regulatory Adjustment
Press, 128 p.
18-26.
Scherer, F.M. (1984), Innovation and Growth: Schumpeterian Perspectives, Cambridge, MA, MIT
Press.
Scholes, Patrick (1950), Capitalism, Socialism and Democracy, 3rd edn, New York, Harper and
Row.
Seldon, Zena A. and Seldon, James R. (1984), ‘Natural Monopolies versus Desirable Monopolies and
Regulation in the Public Interest: Two Quibbles and a Policy Note’, 23(2) Quarterly Journal of
Press, 229 p.
Shepard, W.G. (1965), Economic Performance under Public Ownership, New Haven, CT, Yale
University Press, 161p.
319-327.
133-150.
Sibley, David (1989), ‘Asymmetric Information, Incentives and Price-Cap Regulation’, 20 Rand
Sing, Merrile (1987), ‘Are Combination Gas and Electric Utilities Multiproduct Natural Monopolies?’,
Spence, A. Michael (1975), ‘Monopoly, Quality and Regulation’, 6 Bell Journal of Economics,
Regulation of Natural Monopoly

417-429.


Williams, Stephen F. (1991), ‘Fixing the Rate of Return After Duquesne’, *8 Yale Journal on Regulation*, 159 ff.


Williamson, O.E. (1976), ‘Franchise Bidding for Natural Monopolies - In General and With Respect
Regulation of Natural Monopoly


Other References

