5200

PRICE REGULATION: A (NON-TECHNICAL) OVERVIEW

Janet S. Netz
Department of Economics, Purdue University
© Copyright 1999 Janet S. Netz

Abstract

This chapter surveys the literature on various forms of price regulation. After explaining why natural monopolies should be regulated, I describe rate of return regulation, the traditional approach to regulating prices. I then discuss the resulting problems, primarily the incentive of the firm to overinvest in capital. I discuss peak load pricing and Ramsey-Boiteux pricing as regulatory mechanisms for multi-product firms. I then describe more recent attempts to design regulatory mechanisms that are not subject to the types of disincentives associated with cost-based price regulation. These forms of regulation include price-cap and profit-sharing regulatory mechanisms. I also discuss the regulation of networks, where a regulated firm competes with rivals and supplies them with an essential input, generally access to a network. I conclude by describing how regulation has been practiced.

JEL classification: L51, K23

Keywords: Price Regulation, Price-Cap Regulation, Incentive Regulation

A. Overview

1. Introduction

Economists have long recognized that the market outcome for natural monopolies leaves much to be desired. In particular, price is higher and output is lower than the social optimum. Recognition of this problem, among other issues, has led to a long history of attempts to regulate natural monopolists and to a vast literature discussing the problems of attempts at regulation. (Regulation has been enacted for a variety of other reasons, as well. For example, regulation may be motivated by distribution concerns, as the market outcome for natural monopolies redistributes surplus away from consumers to producers. In other cases regulation may occur at the behest of the firms themselves, who may be seeking protection from too much competition (see Chapter 5000 General Theories of Regulation for a discussion of the political
and economic forces that have led to regulation). The goal of this survey is to lay out the economic (efficiency) problems caused by natural monopolies, the various forms of regulation that have been attempted, and their economic effects.

The chapter is organized as follows. In Section 2, I present the basic issue: that is, why regulate at all? I focus on the natural monopoly problem, though much price regulation occurs outside industries that are natural monopolies. I briefly discuss examples of price regulation that move the outcome towards the social optimum. This section can be skipped by those familiar with the theory of natural monopolies. Perhaps the most widely used early form of price regulation was really rate-of-return regulation, which I describe in Section 3. Though under this model the regulator specifies an allowed rate-of-return, in practice rate-of-return regulation is implemented when the regulator specifies the allowed prices that a regulated firm may charge that the regulator estimates will give the firm the allowed rate-of-return. As summarized in Section 4, Averch and Johnson (1962) show that such an approach to regulation encouraged the utility to overinvest in capital and to expand into other markets in order to increase its rate base and hence increasing the aggregate amount of profits it could earn subject to the regulatory constraint. In Section 5, I describe extensions to the Averch-Johnson model, as well as summarize empirical estimation of these negative effects.

In Section 6, I discuss the multi-product regulated firm. While many of the issues discussed in Part B in the context of a single-product firm apply, the fact that firms produce many products complicates regulation. The largest complication arises because of the existence of common costs; that is, costs that are used to produce more than one product. The issue is discussed in Section 7. In Sections 8 and 9, I discuss two of the most common approaches to regulating a multi-product firm, peak-load pricing and Ramsey-Boiteux pricing. I conclude Part by discussing how the resulting price structure can cause the regulatory regime to be unstable (unsustainable.)

Regulation theory has more recently focused on designing regulatory mechanisms that do not provide the utility with the adverse incentives described above. In Part D, I consider alternative forms of price regulation that fall under the rubric ‘incentive regulation’, focusing on price-cap regulation in Sections 12 and 13 and on other forms of incentive regulation such as profit-sharing in Section 14. In Section 15 I highlight some of the attempts to compare incentive regulation to other forms of regulation, both in terms of their welfare effects and on the feasibility of implementation. Section 16 discusses the few empirical attempts to measure the impact of incentive regulation. In Part E, I discuss one of the most recent developments in regulation: regulation of access pricing in situations where an incumbent firm provides access to an essential facility, generally a physical network, to rivals. I conclude in Part F by discussing price regulation in practice, including the several recent waves of deregulation.
I make no claims that this is an exhaustive survey; indeed, in the interests of brevity, many subject areas will be almost completely omitted. Also, because other chapters in this encyclopedia cover the topics, I will omit discussion of: the theory of regulation, or why different forms of price regulation might be implemented (see Chapter 5000, General Theories of Regulation), as well as regulation in particular sectors (see Chapters 5660, Regulation of the Securities Market; 5700, Insurance Regulation; 5850, Regulation of Banking and Financial Markets; 5900, Transportation; 5930, Telecommunications; and 5940, Public Utilities). I attempt to balance discussion of the theoretical work that has been done in the area with empirical analysis of the effect of various types of price regulation mechanisms.

2. The Problem: Natural Monopolies

As mentioned above, the market does not achieve production and allocative efficiency if technology is such that the industry structure is one of natural monopoly. In essence, a natural monopoly arises if technology and demand are such that it is cheaper for one firm to serve the market than for several firms to serve the market. (See Panzar, 1989, for a complete, technical discussion of the technology that lead to a natural monopoly. The conditions, especially for a multi-product firm, are considerably more complicated than the simple graph I present. (See Chapter 5400, Regulation of Natural Monopolies for a more complete discussion.) To illustrate the problem, consider Figure 1, which illustrates a natural monopoly arising from economies of scale over the relevant range of production for a firm that sells a single-product.

**Figure 1 Market Outcome versus Regulated Outcomes**
The average total cost curve (ATC) is shown to be everywhere declining (and hence the marginal cost curve, MC, is beneath the average total cost curve). Thus, the industry is a natural monopoly; any market structure involving several firms would involve the unnecessary duplication of fixed costs. Assume that the firm (and regulator) can charge only a single price; that is, price discrimination is not allowed. If the firm is not regulated, it will maximize profits by setting marginal revenue equal to marginal cost, leading to price $P_M$ and output $Q_M$. Productive efficiency requires producing at the minimum point of the average total cost curve; clearly the market outcome does not satisfy this condition. Allocative efficiency requires that production occurs where the marginal cost curve crosses the demand curve; again, the market outcome does not satisfy this condition. Relative to the social optimum, social welfare has been reduced by areas $A$, $B$, $C$ and $D$ (termed the deadweight loss).

Price regulation can theoretically lead to the social optimum if regulators specify that price be set equal to $P_E$, where the E subscript denotes ‘efficient’. (However, price regulation is not the only solution. See, for example, Demsetz, 1968, and Williamson, 1976, who discuss how competition for the market, rather than within the market, can lead towards the social optimum.) Then allocative efficiency is met. The outcome has moved towards productive efficiency; pure productive efficiency cannot be achieved simply because demand is not of sufficient magnitude for production to occur at the minimum average total cost curve. However, a firm that charges $P_E$ and produces at $Q_E$ will not generate sufficient revenue to cover costs of production; in particular, the firm will be short by the amount of its fixed cost. Thus, the regulator must alter the regulatory mechanism in order that the firm remains in the market. To ensure that the market is served, the regulator might offer the firm a subsidy equal to its fixed costs. (The general equilibrium effects of the scheme used to raise the subsidy may distort other markets, so that efficiency in the regulated market may obtain at the expense of inefficiency in other markets.) (If a natural monopoly arises via subadditive costs as defined in Panzar, 1989, but without economies of scale over the entire range of production, then marginal cost pricing would not require a subsidy.)

If provision of a subsidy is not politically feasible, the regulator may alternatively specify that the firm charge $P_F$, the price where the average total cost curve crosses the demand curve (the F subscript indicates ‘feasible’). At this price the firm charges the lowest price possible, subject to the constraint that it cover all costs. This regulatory mechanism increases social welfare by areas $A$, $B$ and $C$, relative to the market outcome. Society is still losing area $D$, but this may be acceptable relative to the political cost of providing the firm with a subsidy equal to the firm’s fixed costs.

A variety of other pricing schemes are available that may be feasible and may succeed in achieving the social optimum. In general, these pricing
schemes relax the assumption made above that the firm can charge only a single price. Consider, for example, an alternative scheme that may be available if price discrimination is allowed. The regulator could require the firm to sell at price \( P_f \) to those consumers who are willing and who can afford that price, and charge \( P_e \) to those customers who are willing and who can afford \( P_e \) but not \( P_f \). Revenues earned from customers paying \( P_f \) will cover the variable costs of serving them and all of the fixed costs; revenues earned from customers paying \( P_e \) would cover the variable costs of serving them.

Of course, a practical problem immediately arises. How could the firm classify consumers into the two pricing categories? No consumer would willingly admit that she was willing to pay the higher price. Furthermore, the product must be one that cannot be resold. If it could, then even if the firm could distinguish between types of consumers, the consumers facing the lower price would buy and resell to the consumers facing the higher price. Price discrimination can be adopted only in situations where the firm can identify customers by type and can prevent resale of the product.

Alternatively, consider that a firm may charge different prices for different amounts of the good purchased. A common such scheme, called a two-part tariff, is one where each customer pays a monthly fixed price for access equal to the total fixed costs divided by the total number of customers and then the customer pays an additional fee equal to the marginal cost for each unit consumed. The fixed fee covers the fixed cost of operation and the per unit fee covers variable costs. Since total revenues cover total costs, the firm would not require a subsidy. This pricing scheme is efficient only if consumer surplus, which is the value the consumer places on consuming the product less the cost the consumer must pay, for the consumer with the smallest demand is greater or equal to the fixed price. Otherwise some consumers will exit the market, in which case the scheme does not achieve the social optimum.

Other types of non-linear pricing exist, as well, which may or may not achieve, or have the objective of achieving, efficiency. For example, block pricing, where consumers pay one price for the first block of the product that they use (say, the first 500 kilowatt hours of electricity used), then a different price for the second and subsequent blocks. Standard or declining block pricing refers to declining prices for subsequent blocks, inverted block pricing refers to increasing prices for subsequent blocks. (Block pricing of either kind is a common feature of gas, electricity and water price structures.) If used alone, a declining block pricing scheme optimally charges a price on the first block sufficiently high to cover the fixed costs of operation and charges a price equal to marginal cost on subsequent demand. (This is simply a variation of the scheme described above where some consumers pay \( P_f \) and some pay \( P_e \).) Alternatively, block pricing can be used in conjunction with a fixed fee of the type discussed above. In a declining block pricing scheme, efficiency may be
enhanced over the two-part tariff described above if the two-part tariff has a sufficiently high fixed fee so as to induce some consumers to exit the market. In a declining block pricing scheme with a fixed fee (that is, a multi part tariff), the higher price on initial units can be used to reduce the fixed fee, which increases efficiency by inducing more customers to remain in the market. Inverted block pricing is incompatible with efficiency, and is generally adopted to redistribute income (as in many developing countries; Maddock and Castano, 1991, study the effects of the inverted block structure adopted in Medellin, Colombia for electricity rates on redistribution, the stated objective of the policy, and find that the policy is effective at redistributing income, though of course at a cost to efficiency) or to encourage conservation (as in several municipalities in California for water during the droughts of the 1980s; some succeeded in reducing demand so successfully that they were forced to raise rates in order to generate sufficient revenue to cover costs!). See Train (1991) for a full description of block pricing.

In sum, price regulation of some form can increase social welfare relative to the market outcome. It is theoretically possible to achieve the social optimum, where the marginal cost and demand curves cross, though in practice it may only be possible to move towards the social optimum, not to achieve it.

B. Rate-of-return Regulation and its Problems, or the Unintended Consequences of Regulation

3. Introduction to Rate-of-return Regulation

Price regulation, as practiced in the US, has often involved rate-of-return regulation. Under such regulation, the regulatory agency sets prices in such a way that the utility earns the allowed ('fair') rate-of-return on its investment. While this form of regulation is seemingly simple, and seems as if it would achieve feasible average cost prices, in practice average cost prices are not achieved because the regulatory mechanism gives the firms an incentive not to minimize cost. (In the simple analysis presented above, a competitive rate-of-return to the utility was built into the average total cost curve, as total costs must include an adequate return on investments. Thus, average cost pricing at \( P_r \) is an example of rate-of-return regulation, where the allowed rate-of-return is equal to the competitive rate-of-return.)

Before turning to the effects of regulation, I briefly describe the institutions that were adopted in the US to implement regulation (see Kaserman and Mayo, 1995, for a detailed description of the regulatory process). While historically most other countries chose public ownership as the common method for controlling natural monopolies, the US generally chose to separate the
ownership of the firm from the government agencies that regulated them. (Utilities that serve the majority of demand are investor owned. Typically utilities are granted a franchise (monopoly) over a geographic area, and are subject to regulation.) Because of the movement towards privatization worldwide, the structure of the regulatory bodies may be of more general interest as other countries establish regulatory agencies to oversee newly privatized firms.

In the US, local (retail) electricity and natural gas distribution, local phone service, and private water systems (this last is the most commonly municipally owned utility in the US), are regulated by an independent state public utility commission or PUC. (Long-distance phone rates (until very recently) and wholesale electricity and natural gas transactions are regulated in a similar manner, but by a federal agency, since under the US Constitution interstate transactions fall under the purview of the federal government.) State PUCs consist of from one to seven commissioners, who may be appointed (most commonly) by the governor with legislative approval or elected, depending on the state, and a professional staff.

Prices are set to allow utilities to earn a fair rate-of-return on their capital investment. (The level and structure of prices may also be set to accomplish other goals, such as subsidizing some consumer groups over others.) Rulings by the US Supreme Court mandate that state PUCs provide regulated firms with a fair rate-of-return, whatever the PUCs other goals in setting prices may be. In *Smyth v. Ames* (169 US 466, 1898), the Court stated that ‘The company is entitled to ask for a fair return upon the value of that which it employs for the public convenience’. At the same time, the Court ruled that consumers deserved protection, as well: ‘while the public is entitled to demand ... that no more be extracted from it ... than the services are reasonably worth’. The Supreme Court ruled further in *Federal Power Commission v. Hope Natural Gas Co.* (320 US 591, 1944) that regulated firms are entitled to a ‘just and reasonable’ rate-of-return, so that the firm is able ‘to operate successfully, to maintain its financial integrity, to attract capital, and to compensate its investors for the risks assumed’. In order to set the price structure to generate the fair rate-of-return, the PUC first determines the utility’s revenue requirement, which consists of the sum of operating expenses, current depreciation, taxes, and the allowed rate-of-return times the rate base. The rate base is the value of assets, often measured as the original cost of equipment less accumulated depreciation. (This measure has the advantage of being easy to calculate, though a variety of other methods are used, often involving the replacement cost of capital.) The rate-of-return must be chosen high enough to raise capital, but low enough to prevent the use of market power.

The PUC then forecasts demand and, based on the revenue requirement and the quantity forecast, the agency stipulates the prices that the regulated firm is
allowed to charge. Often the utility proposes a rate structure and, based on information from a ‘test year’, typically the latest year for which the necessary data are available, demonstrate that the price structure would yield the revenue requirement. Ex post, depending on the rate-of-return actually earned by the utility, the firm or consumer groups or the regulatory agency may request or undertake a rate hearing to alter prices if the earned rate-of-return deviates substantially (in either direction) from the allowed rate-of-return. In addition, the staffs of the commissions often periodically monitor the firms’ performance, in which case they may request that the commission hold a new rate hearing.

The method used to gather information and to decide new rates is rather like a court hearing. A variety of groups may give testimony, including the commission staff, the firm, consumer groups and rival firms. In some states consumers are formally represented at regulatory hearings by the Office of Public Utility Counsel. A rate hearing consists of the presentation of evidence to the commission. There are formal rules of procedure à la judicial proceedings. The commission announces a hearing, then publishes the rules of discovery (that is, the rules by which parties are granted access to the factual data of other parties) and a timetable for the hearing. Typically, first the utility is required to file written testimony supporting its rate change. Then, typically 30 to 90 days later, other parties (rivals, consumers, and so on) are allowed to present testimony supporting or opposing the utility’s proposed rate change. Rate hearings often have two phases, the first determining the revenue requirement and the second designing the rates that will allow the utility to earn the revenue requirement.

One area of contention regards costs incurred by the utility. Operating expenses are typically non-controversial. However, as shown in the next section, regulated firms have an incentive to invest in capital beyond the efficient level. Thus, public utility commissions determine whether costs, typically capital costs, are prudent. Typically, a firm that is expanding capacity will construct the new plant, and only once the construction is completed will the rate base change. If the new capital was automatically added to the rate base, then firms may have the incentive to invest in plants that are bigger than necessary in order to inflate the rate base and increase profits. Regulators have the authority to determine the prudence of expenditures on addition to capital. Prudence is determined by whether the capital is necessary to provide the service and whether it is cost effective relative to alternatives. The fact that capital investment may be disallowed encourages the firm to invest wisely. (Most states have to authority to require advance approval of capital budgets; yet even with advance approval, there is no guarantee that the investment will be allowed into the rate base.) The most commonly used example of disallowances involves the building of nuclear power plants. Public Utility Commissions have in fact allowed several utilities to go bankrupt as a result of
disallowances.

4. The Averch-Johnson Model

Averch and Johnson (1962) model the regulated firm as seeking to maximize profits subject to a constraint on the earned rate-of-return, and show that if the regulated rate-of-return is higher than the cost of capital (it must be equal to or higher than the cost of capital if the firm is going to produce at all), then the regulated firm (1) will not minimize costs and (2) may expand into other markets, even if it operates at a loss in these markets, and hence may drive competitive firms out of the market (or discourage them from entering).

The firm will not minimize cost because, so long as the regulated return on capital is higher than the cost of capital, the firm makes a profit for each additional unit of capital. (Indeed, Baron, 1989, points out that the regulated firm would like the price to be set as low as possible, to increase the quantity sold. That gives the firm the opportunity to employ as much capital as possible, and every additional unit of capital employed gives the firm a profit.) This gives the firm an incentive to invest more in capital than is efficient given the actual cost of capital relative to other input prices. Consider, for example, an unregulated firm. This firm will, for each quantity produced, combine labor and capital (assuming these are the only inputs) in such a way that the last dollar spent on labor gives the same output value as the last dollar spent on capital. That is, each input is used until the cost of the last unit employed is equal to the revenues that can be earned with that input’s output. A regulated firm, on the other hand, not only earns revenues by selling the output produced by capital, but also on the capital itself. Hence, each unit of capital provides higher revenue to a regulated firm than to an unregulated firm, encouraging the regulated firm to employ a larger amount of capital than will an unregulated firm. The regulated firm then uses too much capital relative to labor, producing at higher than minimum cost. The degree to which a firm overcapitalizes depends on its ability to substitute between inputs. The more substitution the technology allows, the more inefficient a firm will be.

Now consider a multi-product firm that may operate in several markets, assuming that the regulatory body allocates at least some of the capital used for production in the other markets to the firm’s rate base (as was commonly done in practice, given the problem of allocating common costs; see Section 7). Based on the same logic described above, the firm has an incentive to increase its use of capital, and now may do so in other markets. Suppose that in its initial market the regulated firm has invested in the optimal (from its point of view) amount of capital. The firm now considers investing in additional markets. As an extreme, suppose that this firm is not competitive in another
market: that is, its costs are above rival firms. Then it will lose money on each unit that it sells in the market. However, suppose for each unit of capital that it uses to produce the profit-losing product, it earns revenue because it loosens the rate-of-return constraint. Then the firm will operate in the new market until the extra revenues earned per additional unit of capital are offset by the cost of capital, other inputs, and the loss from sales of the product. That is, a firm will operate in a second, non-competitive market, so long as \((s_1 - r_2)x_2\), where \(s_1\) is the allowed rate-of-return in the original market, \(r_2\) is the cost of capital in the other market, and \(x_2\) is the amount of capital used in the other market that is allowed into the rate base, is less than the losses from operating in the market. (Averch and Johnson, 1962, argue that the descriptive evidence suggests that AT&T behaved along the lines predicted by their model in terms of both overinvesting in capital and moving into other markets.) Thus, not only will the firm’s input use be distorted in the regulated market, but inefficiencies can spill into competitive markets as competitive firms are driven out of business or are prevented from entering markets in which they have lower costs. (Braeutigam and Panzer, 1989, examine this issue more rigorously, arriving at the same conclusion.)

In sum, rate-of-return regulation as practiced provides firms with the incentive to overinvest in capital, both in the regulated market itself, and by inefficient diversification into other markets.

5. Extensions, Reinterpretations and Tests of the Averch-Johnson Model

The Averch-Johnson model led to a variety of work applying the basic idea to other settings. One intellectually unappealing aspect of the Averch-Johnson model is the fact that the regulator does not recognize the firm’s strategic response and hence modify the regulatory mechanism to mitigate the capital disincentives. Baron (1989) recasts the Averch-Johnson model in terms of a mechanism adopted to deal with the regulatory body’s incomplete information on the firm’s costs or on demand. If the regulator could observe the production function and factor prices, the regulator could set price equal to the efficient marginal cost (that is, the marginal cost that arises from the optimal use of capital relative to labor), thus eliminating the production inefficiency. However, suppose that the regulator does not observe the production function and/or factor prices. Then in essence, the regulator adjusts price as a function of the cost incurred by the firm. In Baron’s view, the Averch-Johnson outcome is plausible in a setting where the regulator and the firm have different information about costs and demand or the regulator cannot observe all actions of the firm. In such a setting, the Averch-Johnson approach could arise endogenously. Along similar lines, Woroch (1984) analyzes investment by firms using a non-cooperative model of the firm regulator relationship, and
allows the regulatory mechanism to arise endogenously as well. He allows firms to invest strategically. Woroch finds that achieving ‘credibility’ of regulation, which requires that the regulator issue a welfare maximizing response to the capital investment, may lead to an outcome that is indistinguishable from the Averch-Johnson model. In other words, the Averch-Johnson model may be recast in different, perhaps more plausible, frameworks, but the disincentives caused by regulation may still arise.

Bailey and Coleman (1971), Baumol and Klevorick (1970), Bawa and Sibley (1980), Davis (1973), Klevorick (1973) and Logan, Masson and Reynolds (1989) extend the Averch-Johnson model to incorporate regulatory lag. These authors recognize that, in practice, rate hearings occur with a lag after a change in cost or demand. Under regulatory lag, a firm can appropriate increased profits from cost reduction for the time period between the innovation and the rate hearing (and likewise, the firm absorbs any losses). This reintroduces to the firm the incentive to reduce costs, though the incentive is not as large as it would be if the firm appropriates gains from cost reductions forever.

Encinosa and Sappington (1995) and Lyon (1991, 1992) examine a firm’s incentive to over- or under-invest in capital under a situation where regulators review a firm’s capital investments and can (and do) disallow some capital expenditures. The Averch-Johnson model was based on a more certain environment where the firm knows that it will earn the allowed rate-of-return on its entire capital investment. More recently, regulatory agencies have become tougher in allowing capital expenditures into the rate base. It has been typically accepted that regulatory hindsight, especially if regulators punish bad ex post outcomes rather than bad ex ante decisions, will lead to underinvestment in capital. Lyon (1991), on the contrary, finds that hindsight review can be used to reduce the firm’s tendency to build large, risky projects, moving the firm closer to the cost-minimizing input choice. In addition, Encinosa and Sappington (1995) find that rewards as well as penalties are optimal. Lyon (1992) examines prudency reviews in the context of contracts for variable inputs, and finds that prudency reviews will cause a firm to increase its capital stock and to rely more heavily on spot market purchases for variable inputs; while the firm earns lower profits, the welfare effects on consumers are ambiguous.

The Averch-Johnson model has also been included in a dynamic setting including adjustment costs of investment. For example, El-Hodiri and Takayama (1981) show that the long-run level of capital stock increases when a rate-of-return constraint is imposed. Moretto (1989) adjusts the Averch-Johnson model in two dimensions: to incorporate an adjustment cost approach to investment and to allow for uncertainty over the revenue function. The Averch-Johnson effect may not obtain; it holds only when the rate of growth of the marginal adjustment cost is concave. In contrast to El-Hodiri and Takayama (1981) and Appelbaum and Harris (1982) but along the lines of
Moretto (1989), Dechert (1984) adopts a dynamic adjustment-cost model and finds that underinvestment may obtain, in contrast to the Averch-Johnson overinvestment result. The factor driving is result is whether a firm is operating in the increasing return or decreasing return portion of average cost. (This holds for the static Averch-Johnson model as well.) Dechert (1984) argues that natural monopolies arise under economies of scale, in which case profits are not concave, at least not over the entire input range; in contrast, other Averch-Johnson models assume concavity of the profit function. (Note, however, that natural monopolies can obtain outside the increasing returns portion of the average total cost curve; see Panzar, 1989, for an explanation.) The importance of whether costs are increasing or decreasing may explain the mixed empirical evidence regarding the Averch-Johnson model.

A fair bit of empirical work has been done trying to test the Averch-Johnson model, mostly focusing on electricity production. The results of these tests have been mixed. Nemoto, Nakanishi and Madano (1993) find that seven of nine Japanese electric utilities do significantly overinvest, while Tawada and Katayama (1990) find much weaker support for the Averch-Johnson effect in the same industry. Indeed, they find that production was efficient in some time periods, despite the form of regulation. In the US, Courville (1974), Hayashi and Trapani (1976) and Spann (1974) all find support for the Averch-Johnson effect in electric utilities. Hayashi and Trapani (1976) also find that tightening regulation increases the distortion. Boyes (1976), on the other hand, does not find support for the Averch-Johnson effect, also using data on electric utilities in the US. Hsu and Chen (1990) find empirical evidence of the presence of an Averch-Johnson effect for the Taiwan Power Company. Oum and Zhang (1995) find empirical support for the hypothesis that the introduction of competition in the US telephone industry induces incumbents subject to rate-of-return regulation to use capital closer to the optimal level, reducing the Averch-Johnson effect.

Thus, theoretical models support the Averch-Johnson findings of a tendency towards overinvestment in capital by firms subject to a rate-of-return constraint, though introducing some modifications of the framework in order to make it more realistic, such as regulatory lag and cost disallowances, mitigates the overinvestment effect to some extent. There is somewhat more empirical support for the Averch-Johnson effect than against it. As I will discuss in Part D, the form of regulation has changed a great deal, much of it in response to the debate over the Averch-Johnson overcapitalization effect in the face of rate-of-return regulation.
C. The Multi-product Regulated Firm

6. Introduction to Multi-product Issues

The discussion thus far has focused on a single-product firm (except during the discussion that a regulated firm has an incentive to enter other markets in order to increase its rate base). In reality, of course, most regulated firms produce a variety of products. Even in the case of a single commodity, say electricity, it is more proper to think of the firm providing electricity during certain time periods. In other words, the market for electricity during the day is separate from the market for electricity during the evening and night-time. Before discussing various price mechanisms for regulating the multi-product firm, an important concept must be addressed: how to allocate common costs, that is, costs common to production of several products, to individual products? (If the goal of regulation is to maximize social welfare, that is, to achieve allocative and productive efficiency, separating costs across services is meaningless. It is generally done by regulators, however, when they have an interest in the distribution of surplus across groups; that is, regulators often desire to determine whether a given customer group is covering the costs of servicing them.) After describing the problem, I discuss two of the more commonly used regulatory mechanisms, peak-load pricing and Ramsey-Boiteux pricing.

7. The Common Cost Problem

Consider an example, following Braeutigam (1989), where a firm produces two products with constant marginal costs of $m_1$ and $m_2$ and has total fixed costs $F$, giving a simple total cost function of $TC = F + m_1 y_1 + m_2 y_2$. Analogously to the discussion of a single-product firm, the social optimum requires that each product be priced at its marginal cost; however, the firm will then not earn sufficient revenues to cover costs. Recall that a simple solution in the single-product case was to require that the firm price at average cost. The problem here is that the calculation of the average cost for each product necessitates a division of the fixed costs across the two products. While infinitely many ways exist to divide the fixed costs, some will lead to higher levels of social welfare than others. Thus the more interesting problem is how to allocate fixed costs in the optimal manner.

In practice, a variety of weighting schemes have been used, including weighting by gross revenues, by physical output levels, or by directly attributable costs. (Directly attributable costs typically include all variable costs, and may also include some portion of fixed costs, if some fixed costs are associated with the production of only one of the products.) Alternatively,
sometimes the Allais rule is used, where marginal cost proportional mark-ups are used, with the mark-up being determined in such a way as to raise sufficient revenue that the firm breaks even. Attributing costs in this manner is typically referred to as fully distributed cost. In practice, once the common cost is assigned to the different outputs, prices are set so that revenues from each output cover costs.

There are several problems with this approach to dividing costs for a multi-product firm, including the fact that such division is inherently arbitrary and, from an economic point of view, not very sensible. As summarized by Laffont and Tirole (1996), the drawbacks include the fact that, since it is cost-based, firms do not have a strong incentive to minimize costs; that the price structure is inappropriate as prices are too low for inelastic demand services and too high for elastic demand services (see the discussion of Ramsey-Boiteux pricing in Section 9); that it does not use non-linear pricing; and that some forms encourage inefficient entry.

Economists have developed several pricing approaches to maximize social welfare given the constraint that the multi-product firm must break even. These pricing schemes do not require allocation of common costs. (Ex post one could calculate the proportion of the common cost that is paid for by each output; however, ex ante the only necessary information needed is the total amount of fixed cost.) I discuss two of the most common methods: peak-load pricing and Ramsey-Boiteux pricing.

8. Peak-load Pricing

Peak-load pricing is a particular regulatory mechanism that is appropriate when the firm provides a non-storable good during different time periods and is constrained to choose a single capacity level. (See Crew, Fernando and Kleindorfer, 1995, for a survey on peak-load pricing theory.) Formulating prices for different time periods relates to the common cost problem: optimal prices depend on how much of the capacity cost, a common cost, is allocated to each time period. Consider the following simple model. A production period (a day or year) is divided into $T$ equal increments. During each period $t$, $x_t$ units of a variable input are used. Let $K$ denote the total amount of capital used. $K$ is the common cost, the cost of the production facility that operates in all periods. Let $y_t = f(x_t, K)$ represent the production function, and let $P_t = P_t(y_t)$ denote the inverse demand curve (the notation and example are based on Braeutigam, 1989).

Steiner (1957) assumed that the production function was Leontief; in other words, $y_t = \min(x_t, K)$, where, for notational simplicity, I have assumed one unit of capital and one unit of the variable input produce one unit of output. (The qualitative results hold for an fixed proportion of capital to the variable
input. Let the cost of the variable input be given by \( b \), and let \( \beta \) be the rental cost of capital. Assuming that the firm is constrained by the regulatory authority to meet demand in all periods, it will choose capacity to equal the maximum amount demanded. Then the total cost function can be written as

\[
TC = b \sum_{t=1}^{T} y_t + \beta \max y_t,
\]

In other words, total cost is equal to the cost of the variable input times the total amount of the variable input used in all time periods plus the rental cost of capital times the maximum output produced in any period.

The social optimum requires that price be set equal to marginal cost in each period. Based on the total cost function, this means that the price in every non-peak period is optimally set equal to the cost of the variable input, \( P_{\text{non-peak}} = b \).

(Marginal cost in a non-peak period is given by \( dTC/dy = b \), the derivative of the first term in the cost function, since the second term does not contain output in any period except the peak period.) Price in the peak period is set equal to the cost of the variable input and the cost of capacity, \( P_{\text{peak}} = b + \beta \). In other words, consumers during the peak period are responsible for both their variable costs and total capacity costs. Given that the technology exhibits constant returns to scale, peak-load pricing leads not only to the social optimum, but also allows the firm to earn sufficient revenues to cover total costs, obviating the need for a subsidy.

However, Panzar (1976) shows that this result is not robust to changes in technology. Suppose that the production function is given in a general form as \( y_t = f(x_t, K) \), so that the firm can substitute to some extent between the variable and the capital inputs. The production function is assumed to be subject to diminishing returns (that is, increases in either input increase output, but at a declining rate). Let the total variable cost function be written as \( V(y_t, b, K) \). This function gives the minimum total variable costs required to produce output \( y_t \), given input prices \( b \) and capacity \( K \). First, pricing in each period is given by \( P_t = M/V(y_t) \). This condition indicates that price in each period is equal to the marginal variable cost. In this case, prices in off peak periods will not be equal since marginal cost varies with output. For example, suppose that output in period 2 is larger than output in period 1. Given the assumptions on the production function that an increase in inputs increases output at a declining rate, the marginal variable cost will be higher in period 2 than in period 1, and therefore so will the price in period 2.

Second, optimal pricing requires that \( S \sum_{t=1}^{T} M/V(y_t) = \beta \). This condition shows that the firm will choose capital such that the total variable costs saved equals the cost of employing the last unit of capital. The model using Leontief production did not have a second condition, because the firm had no choice in capital. The firm was required to produce enough to satisfy demand in each
period (a regulatory constraint), and technology was such that satisfying the regulatory constraint dictated the amount of capital used. In the production function used here, the firm can substitute between the variable input and the capital input, choosing the cheapest ratio of the two, depending on their relative prices. Thus, while peak-load optimal prices can be derived from a more flexible production function, it no longer remains true that only the peak consumers pay for capacity costs. However, it remains true that peak-load pricing raises sufficient revenues to cover costs.

The existence of uncertainty over demand or supply can complicate the application of peak-load pricing. Koschat, Srinagesh and Uhler (1995) derive peak-load prices when demand in different periods is stochastic and test the implications of the theory using data on local telephone service. They find that ignoring uncertainty in demand can lead to inefficient prices and inefficient industry-wide capacity choices. Perhaps most interestingly, they find that spot pricing of phone calls, so that prices are high during actual congestion rather than during times of expected congestion, can be considerably more efficient. Kleindorfer and Fernando (1993) also consider peak-load pricing under uncertainty in demand, but add to the analysis by considering uncertainty in supply as well, focusing particularly on electricity (where uncertainty in supply is a common problem). In this setting, the regulatory agency must account for a variety of costs that arise when supply and demand do not clear.

9. Ramsey-Boiteux Pricing

The peak-load pricing mechanism described above, by assuming constant returns to scale, led to prices that were not only socially optimal, but under which revenues covered costs. Thus, peak-load prices are efficient and feasible. However, without constant returns to scale, marginal cost pricing will not raise sufficient revenue that the firm will cover total costs. As an alternative, Ramsey-Boiteux prices can be used to satisfy the constraints that the firm charge a single price for each product and that the firm raises enough revenue to cover total costs. (Ramsey, 1927, introduced this pricing scheme as an optimal form of taxation that raised the required revenues at the lowest deadweight loss cost. Boiteux, 1956, applied the idea to the determination of regulated prices.)

Assuming the demands for the different products are independent, Ramsey-Boiteux prices are the solution to the following constrained maximization problem,

\[
\max_p \left( CS(y) + p \cdot y - C(y, w) \right) \\
\text{s.t. } p \cdot y - C(y, w) \geq 0,
\]
where $CS(y)$ is consumer surplus as a function of the entire output vector $y$; $p$ is the price vector; and $C(y, w)$ is total cost as a function of outputs and the vector of input prices $w$. The social planner (regulator) maximizes social welfare, defined as the sum of consumer surplus and firm profits, subject to the constraint that profits be nonnegative.

Solving the maximization problem leads to the following relationship:

$$
\left[ \frac{p_i & C_{y_i}}{p_i} \right] e_{ii} \cdot \left[ \frac{p_j & C_{y_j}}{p_j} \right] e_{jj}, \quad i, j,
$$

where $C_{y_i}$ is the marginal cost of product $i$, $p_i$ is the price of product $i$, and $e_{ii}$ is the elasticity of demand for product $i$. This relationship, one of several formulations of the Ramsey-Boiteux pricing rule, shows that optimal prices are such that the price-cost margin multiplied by the product’s elasticity of demand be equal across all products produced by the firm. Thus, the price of the more inelastic good is higher than the price of the elastic good.

How is it that prices set according to this pricing rule maximize social welfare, subject to the constraint that firms earn zero economic profits? (Dierker, 1991, formally proves that Ramsey-Boiteux prices are second best prices.) First, note that maximizing social welfare is equivalent to minimizing the deadweight loss. Then the optimality of Ramsey-Boiteux pricing can best be explained graphically. Figure 2 illustrates two related markets. In the upper panel, the firm is constrained to set the same price in each market. Price $P$ is chosen such that the areas $A$ and $B$ equal the fixed costs faced by the firm, so that the firm just breaks even. In the lower panel, the firm is allowed to choose separate prices for each market, and the areas $C$ and $D$ also add up to fixed costs. The shaded areas in both panels are the deadweight loss arising from the fact that price deviates from marginal cost. The two shaded areas in the lower panel are in sum smaller than the two shaded areas in the upper panel, reflecting the fact that the prices chosen in the lower panel minimize the deadweight loss. Intuitively, this comes about because demand in the left panels is inelastic; thus, price can be increased in this market with a relatively small deadweight loss, since as price rises, demand does not decline too much. On the other hand, increasing price in the right panels where demand is elastic induces a large decline in quantity, and hence a large deadweight loss.
The description of Ramsey-Boiteux pricing presented above assumes that demand for the two products be independent. This may be an overly strong assumption, but the approach can be extended to the case where demands for the products are interdependent; see, for example, Dreze (1964), Rohlfs (1974) and Zajac (1974b). In this case, one must take into account the cross partial derivatives; that is, the impact of the price of one good on the quantity demanded of the other good(s). Dreze (1964) derives the Ramsey-Boiteux pricing rule when demand is not interdependent, and arrives at a pricing rule similar in form to that above. More precisely,

\[
\left[ \frac{p_i \& C_{y_i}}{p_i} \right] (e_i \& e_{ij})^i, \left[ \frac{p_j \& C_{y_j}}{p_j} \right] (e_j \& e_{ij}), \quad \forall i, j,
\]

Figure 2 Deadweight Loss with a Single Price (top panel)
Deadweight Loss with Ramsey-Boiteux Prices (lower panel)
where $e_{ij}$ is the elasticity of good $i$ with respect to the price of good $j$. In essence the net elasticity, $e_i - e_{ji}$, gives the effect of $i$’s price on its own demand less the effect on demand for the other product. (Note that this form of the Ramsey-Boiteux pricing rule reduces to the original form if the cross elasticities are zero. An alternative formulation of the Ramsey-Boiteux rule when demand is interdependent involves ‘superelasticities’, which account for substitution and complementarity among goods. See, for example, Laffont and Tirole, 1993.) Ramsey-Boiteux prices have also been derived in more complicated settings. For example, Horowitz, Seeto and Woo (1996) assume that costs are unknown to both the regulator and the firm. Passmore (1984) assumes demand is uncertain and that the firm chooses capacity simultaneously with prices. Berry (1992) incorporates risk in marginal cost. Braeutigam (1979b) incorporates competition in some products produced by the firm. Brock and Dechert (1983) and Braeutigam (1983) extend Ramsey-Boiteux pricing to a dynamic setting.

The previous work has assumed that the regulator sets prices. But a Ramsey-Boiteux price structure may endogenously arise when firms choose prices, depending on the regulatory regime. Logan, Masson and Reynolds (1989) develop a regulatory mechanism under which the firm is allowed to choose prices subject to a rate-of-return constraint. Regulators review performance, and there is a higher probability that regulators will request a new rate hearing as the firm’s return deviates from the constrained rate. If regulators do not review negative returns too quickly relative to positive returns, the model predicts that the firm’s prices will converge to Ramsey-Boiteux prices and that productive efficiency will obtain. The model indicates that relatively simple regulatory mechanisms that do not require much information lead to Ramsey-Boiteux efficient prices. (As discussed in Section 12, price-cap regulation, in some circumstances, also leads firms to choose prices that converge to Ramsey-Boiteux prices.)

A number of empirical studies have been undertaken to compare regulated prices to Ramsey-Boiteux prices; that is, to determine whether regulated prices are set at the optimal (from a constrained efficiency maximization point of view) level. Matsukawa, Madono and Nakashima (1993) examine pricing by Japanese electric utilities. They find that the prices do not satisfy the Ramsey-Boiteux pricing rule, and that moving towards Ramsey-Boiteux prices may require a large increase in residential electricity rates and a slight decrease in industrial electricity rates. Liu (1993) examines local telephone pricing in Taiwan and also finds that prices do not satisfy the Ramsey-Boiteux pricing rule. In the US, DeLorme, Kamerschen and Thompson (1992) find that electricity generated via nuclear power is not priced according to the Ramsey-Boiteux rule either, but rather that prices respond to political influences.

To my knowledge, no empirical study has found regulated prices that conform to the Ramsey-Boiteux ideal. Perhaps the most obvious explanation
is that regulators have a goal that is different from minimization of the deadweight loss subject to a break even constraint. For example, Sheehan (1991b) argues vehemently that Ramsey-Boiteux pricing is inappropriate for telecommunications regulation, in part because it does not protect ‘captive’ consumers from the use of market power. Friedlaender (1992) considers regulation of a firm that serves two markets, one of which is competitive so that prices are at marginal cost, and one of which is ‘captive’ so that consumers in this market have no alternative supplier. Then the captive sector bears the entire revenue burden. She applies her model to freight transportation by railroad, and shows that the application of Ramsey-Boiteux pricing would cause the price to captive coal shippers to rise to ‘socially unacceptable levels’ since they would be forced to bear the entire revenue requirement. Ashraf and Sabih (1992) extend the Ramsey-Boiteux model to incorporate ‘life-line’ rates (that is, the idea that all households should be able to afford a minimum amount of telephone service) and still find empirically that the pricing structure of electricity in Pakistan deviates from the amended Ramsey-Boiteux prices.

Thus, while Ramsey-Boiteux prices have a number of desirable characteristics from a strict efficiency point of view, in practice efficiency and other concerns lead regulators to deviate from them, often to a substantial degree.

10. Non-linear Pricing and Sustainability

Non-linear pricing (including two-part tariffs, block pricing, peak-load pricing, and Ramsey pricing) lead to issues of sustainability. That is, given non-linear pricing, can the regulated monopoly continue in existence? Two issues in particular arise: the likelihood of cream-skimming and of bypass.

Cream-skimming refers to entry of firms into the high-profit markets served by the regulated monopolist. Without entry, the profits earned in one market are used to ‘subsidize’ prices in another market, either for efficiency reasons as in peak-load and Ramsey-Boiteux pricing or for fairness reasons as in business phone rates subsidizing residential rates. An entrant observes profits to be made in the high-price market, and may be induced to enter that market only. (This is named cream skimming because the entry skims the cream (high profits) and leaves the milk (low profits).) For example, in the US, MCI was interested only in entering the long distance market, from which AT&T was obtaining revenues to keep prices for local calls substantially below cost. (Note that such entry may not be efficient. Price in the high-profit market is above the incumbent’s marginal cost, allowing firms with higher marginal cost, but a marginal cost less than the regulated price, to enter the high-profit market.) In addition, high rates in some markets or during some time periods may
induce consumers, typically large commercial consumers, to bypass the regulated monopoly and obtain the product elsewhere. (Cream skimming is often characterized as bypass, since, for example, MCI was able to bypass the long-distance network owned by AT&T.) Again, the regulated firm sees a reduction in revenues in the high price market that were used to subsidize rates in other markets.

The existence of cream-skimming and bypass combine to make the regulated monopoly less financially viable, since they are deprived of a large source of revenue yet often continue to be obligated by regulators to provide service to at least some segments of the market at low prices. Regulators in such a situation must consider several questions. For example, how much bypass should be allowed? (Recall that regulators often control entry in these situations.) Which firms should be allowed to enter? How should prices be restructured in the face of entry and bypass?

The theoretical literature discussing these issues is somewhat limited. Einhorn (1987) considers a case where the potential entrants have access to a bypass technology to which the regulated monopolist does not have access. He shows that the optimal pricing structure involves the monopolist charging a usage price below marginal cost to the largest users, that is, those users most likely to bypass the monopolist. (These consumers still generate net revenue to the firm since they also pay a fixed fee.) Curien, Jullien and Rey (1998) find, similarly to the earlier work, that smaller users are charged a price above marginal cost while larger users are charged a price less than marginal cost. When bypass is regulated, too little bypass occurs, while competitive bypass (that is, free entry) leads to too much bypass. In any case, distributional effects occur, with large users gaining at the expense of either small users or taxpayers, depending on how the loss of revenue is financed (by raising prices to small users or subsidizing the monopoly with tax dollars).

Laffont and Tirole (1990a) consider optimal pricing strategies given the possibility of bypass under conditions of asymmetric information. They assume two sources of asymmetric information: that the firm cannot distinguish between high and low demand customers and that the firm has superior knowledge about its technology relative to the regulator. (Einhorn, 1990b, considers asymmetric information only in the latter dimension.) The first assumption prevents a charge to high-demand customers below marginal cost, as arises in Einhorn (1987) and Curien, Jullien and Rey (1998), because the customers cannot be typed, and hence third degree price discrimination is not an option. Laffont and Tirole (1990a) find that bypass should be prevented when the regulated monopolist is efficient, and support the finding of Einhorn (1987) and Curien, Jullien and Rey (1998) that charging high demand consumers a price below marginal cost is optimal since keeping these customers contributes towards covering fixed costs through the fixed fee of a two-part tariff. They also find that low-demand consumers may benefit from
the possibility of bypass, if bypass constrains the monopolist (that is, the monopolist reduces prices in order to deter entry).

Some empirical work has been done that is related to cream-skimming and bypass. One common theme has been regarding natural gas prices in regional markets. For example, De Vany and Walls (1996) and Walls (1994) find that local bypass and access to pipeline transportation lead to natural gas prices that move with field prices, as one would expect due to arbitrage. Parsons and Ward (1995) find that larger long-distance phone companies are more likely to bypass the local phone companies switched services than are smaller long-distance companies. Finally, Donald and Sappington (1995) find that as bypass activity increases, regulators are less likely to adopt incentive regulation (which is discussed in the next section), since bypass activity reduces the gains a regulated monopolist can attain by investing in cost reduction.

This concludes the discussion of regulation as traditionally practiced in the US I now turn to the more recent advances in price regulation, where the UK has led the way as it moved from a system of largely nationalized firms to regulated, privatized companies.

D. Incentive Regulation

11. Introduction to Incentive Regulation

In large part, the problems arising in the Averch-Johnson model are due to the fact that the regulated firms do not have an incentive to operate at minimum cost. (As discussed in Section A5, a variety of circumstances exist to mitigate the Averch-Johnson effect, such as regulatory lag and prudence reviews. However, while the deviation from cost-minimization may be reduced, in most if not all formulations of rate-of-return regulation it is not eliminated.) Prices are set such that revenues cover operating costs plus the allowed rate-of-return on investment. Thus, any cost savings will be passed on to consumers via a lower regulated price. Similarly, the firm need not worry about rising costs, because those costs will be covered with a higher regulated price. Part of the motivation behind incentive regulation is to provide the firm with the motivation to behave in a manner more consistent with the social optimum (Crew and Kleindorfer, 1996, present an overview of incentive regulation in the UK and the US).

One of the earliest forms of incentive regulation, price-cap regulation, was designed to eliminate the cost disincentives other forms of price regulation had caused; an additional benefit arose because price-cap regulation reduces regulatory administrative costs (at least in theory). Thus, any cost savings are, at least in part, retained by the firm, reinstating the incentive of the firm to minimize cost. (On the other hand, because firms have more price flexibility,
they also have a greater opportunity to exercise market power.) In addition, in a pure price-cap regime, the regulated market is effectively decoupled from other markets, eliminating the inefficient incentives firms subject to rate-of-return regulation face in regards to entering markets in which they are not competitive. Finally, price-cap regulation may have added benefits in industries where technology is changing rapidly. Maintaining rate-of-return regulation in such situations is administratively difficult, as the regulator must convene more frequent rate hearings and has less past information on costs on which to base pricing decisions. Price-cap regulation requires less information regarding technological change to implement. (This may explain why telecommunications regulation adopted price-cap regulation more quickly than other industries. Einhorn, 1991a; Mitchell and Vogelsang, 1997; and Sappington and Weisman, 1996a provide comprehensive treatment of incentive regulation in the telecommunications industry.)

Price-cap regulation was adopted in the United Kingdom in 1984 to regulate the recently privatized telecommunications industry; the United Kingdom now applies this form of regulation to gas, electricity, water, as well as telecommunications. (Beesley and Littlechild, 1989, and Armstrong, Cowan and Vickers, 1994 describe incentive regulation in the United Kingdom.) It has also been adopted for use in the US, primarily in telecommunications, at the state and federal level. By 1993, thirty states had adopted a form of price-cap regulation for some industries, though generally the mechanism included an aspect of profit-sharing, as discussed in Section 14. The FCC plan divided AT&T’s products among three baskets, each subject to a price-cap, in order to reduce or eliminate cross subsidization between the baskets. Several other countries have also adopted this form of regulation, including Australia, France, Hong Kong, the Netherlands, Mexico, Germany, Sweden and Denmark.

To give an idea of how price-cap regulation works in general, consider the following scenario. The regulatory agency chooses two values in setting the price-cap. The initial price-cap is chosen, P₀; then the regulator chooses an adjustment term, x, so that subsequent price-caps are equal to P₀! x. The adjustment term may occur at any point in time specified by the regulator; often it is an annual adjustment. During the time the price-cap is in effect, the regulated firm may either choose a price below or equal to the price-cap. (Note that the firm is granted a high degree of pricing flexibility; in particular, the firm typically acquires substantial flexibility over the structure of prices. Thus, price-cap regulated firms are relatively free to abandon cross subsidies that may have been imposed under rate-of-return regulation. In the US, subsidies to, for example, rural areas, continue in a different form: telecommunications firms are taxed, the proceeds used for direct subsidization, rather than accomplishing subsidization via the price structure.) The firm can request a rate hearing, which it might do, for example, if its costs were higher than the
price-cap. At some specified point in the future, either price-cap regulation
continues with a hearing to establish a new price-cap, or the regulatory
mechanism reverts to a rate-of-return approach.

12. Price-cap Regulation

(Vogelsang and Finsinger, 1979, derive a regulatory mechanism that is similar
to price-cap regulation.) One of the key assumptions is that marginal cost is a
function of an investment in cost reduction, denoted as $e$ (for effort). A lag
occurs between the investment in cost reduction and its realization, and the
effect of $e$ on cost reduction may be certain or uncertain, so long as higher
effort increases the likelihood that marginal cost will fall. Let $C$ denote the
firm’s original marginal cost; the price-cap is $P_0$ and is set below $C$; and let
$M(C)$ denote the unconstrained monopoly price for a marginal cost of $C$. Then
the price actually charged by the regulated firm, which is a function of the
price-cap and of marginal cost and is denoted as $B(P_0, x, C)$, is equal to max
$\{C, \min\{P_0, x, M(C)\}\}$. If costs are above the price-cap, the firm asks the
regulator for a rate hearing and essentially chooses to revert to rate-of-return
regulation. If costs are below the price-cap, but not so low that the price-cap
is not binding, then the firm prices at the cap. Finally, if costs are low enough
that the monopoly price is less than the price-cap, $M(C) < P_0$, then the firm
prices at the monopoly price. The firm’s maximization problem with respect
to investments in cost reduction can be written as
$$\max_x \pi(P_0 - x, e) = \left[ B\left( P_0 - x, C(e) \right) - C(e) \right] \cdot D\left[ B\left( P_0 - x, C(e) \right) \right] - e,$$
where $D[B]$ is the demand curve as a function of the price charged by the
regulated firm.

Cabral and Riordan (1989) derive several illustrative results. First, in the
certainty case (that is, when the firm knows by how much marginal cost will
be reduced for a given investment $e$), they show that if the price-cap is too tight,
then the firm has no incentive to invest in cost reduction because the return to
innovation is too small relative to the cost. Essentially the firm opts for rate-of-
return regulation. Second, they show that if the price-cap is binding, then an
increase in $x$, that is a tighter price-cap, marginally increases the investment in
cost reduction. The increased incentive is caused by the fact that the tighter the
price-cap, the more units the firm is selling; the aggregate increase in profits
for a decline in marginal cost is higher than it would be for the same decline in
marginal cost at a higher price-cap, since at the higher price-cap, a smaller
quantity would be sold. In the case of uncertainty (that is, when the firm does
not know the cost that will result from a given level of effort $e$), the results are
not as tight. They are able to show that for low values of $x$, the optimal choice of effort increases as does $x$ (that is, a tighter price-cap leads to more cost reduction). In summary, Cabral and Riordan (1989) show that a firm regulated under a price-cap mechanism will have an incentive to invest in cost reduction, depending on the level of the price-cap. However, if the price-cap is too low, the firm may have no incentive to invest in cost reduction.

Clemenz (1991) extends the work of Cabral and Riordan (1989) by deriving the effect of price-cap regulation on consumers and by deriving the socially optimal price-cap. (Clemenz, 1991, notes that the impact on consumers is important since, unlike economists, regulators may place a higher weight on consumer welfare than producer welfare.) Cabral and Riordan’s (1989) basic results are confirmed: in a finite horizon model, price-cap regulation gives the firm larger incentives to reduce cost than does rate-of-return regulation and achieves higher social welfare. These conclusions are even stronger in an infinite horizon setting. Clemenz (1991) also shows that price-cap regulation is capable not only of raising welfare in general but also in increasing consumer surplus.

Sibley (1989) and Lewis and Sappington (1989) derive the optimal regulatory policy in a given environment characterized by asymmetric information; the optimal regulatory mechanism is similar to price-cap regulation. Sibley (1989) assumes the firm has full information about technology and demand while regulators have no information; he also assumes that regulators can observe lagged expenditures by the firm, prices, and outputs. Sibley’s (1989) scheme is derived such that it leads to efficient pricing and input usage; the optimal scheme is remarkably similar to a price-cap regime. Lewis and Sappington (1989) present a model in which the regulated firm has private information about its capabilities and cost-reducing activities. The endogenously-determined optimal regulatory policy includes a form of price-cap regulation as a component (firms are offered a choice between price-cap regulation and a regulatory regime that shares gains with consumers). They suggest that while price-cap regulation may be superior in some environments, it may not be in others, and hence should be considered as part of the optimal regulatory regime.

A number of papers have considered the effect of price-cap regulation on the structure of prices. In the multi-product setting, price-cap regulation is generally implemented by imposing a price index on a basket of goods, giving the firm the freedom to adjust the price structure. A possible additional benefit arising from price-cap regulation is that the firm will choose prices across products in a way that will lead prices to have a Ramsey-Boiteux structure, which allows the firm to cover costs in the most efficient manner (that is, at the smallest deadweight loss). Bradley and Price (1988) and Vogelsang (1989) have demonstrated this result. Others, however, have shown that the result is not general. For example, Neu (1993) finds that prices approach a Ramsey-
Boiteux structure only in certain situations, such as under static conditions. By example he shows that under a number of parameter conditions relating to differential growth in demand, differential price elasticities, and different shares across services, prices will diverge from the Ramsey-Boiteux structure. Fraser (1995) finds that cost increases for a firm’s products can cause prices to diverge from the Ramsey-Boiteux structure if the cost increases or demand elasticities are not uniform. However, if additional cost changes do not occur and the previous period outputs are used as weights in the price-cap, then prices will converge to the Ramsey-Boiteux structure. He concludes that only when cost changes are recurring will price diverge from the Ramsey-Boiteux structure on a continuing basis. Bradley (1993) points out that, since the definition of commodity bundles is largely endogenous (for example, should day time rates end at 4pm, 5pm, or 6pm?) and under the influence if not sole discretion of the regulated firms, the definitions may be manipulated in such a way that Laspeyres quantity based price-caps may very well not lead to an efficient Ramsey-Boiteux price structure. (The role of indices is discussed in the next section.)

Thus, price-cap regulation can give firms a larger incentive to invest in cost reduction than does rate-of-return regulation and the price structure may approach the second best Ramsey-Boiteux prices, and there is some evidence that price-cap regulation is, or can be, part of the optimal regulatory mechanism.

13. Problems and Extensions of Price-cap Regulation

While the discussion above casts a positive light on price-cap regulation and its effects on innovation and efficiency, subsequent theoretical work and experience from the implementation of price-cap regulation suggest that it is not the panacea that it once seemed. Problems arise in terms of the proper calculation of the price index, the optimal pricing structure, the impact on quality, and the end game and renegotiation problems.

Typically the price-cap is implemented as requiring that a price index of the different products be below the price-cap; at issue are the weights used to form the price index. Law (1993) shows that Laspeyres index price-cap regulation under revenue weights causes the firm to price in order to manipulate the weights in such a way that reduces consumer and increases producer welfare in the first period and may reduce consumer welfare in the second period. Foreman (1995) extends the analysis to indicate when and why such weight manipulation incurs. Suppose the price-cap is set in the following way...
422

Price Regulation: A (Non-Technical) Overview

\[
\left( \frac{p_t^i}{p_{t+1}^i} \right) \times \left( \frac{p_{t+1}^i x_{t+1}^i}{p_{t+1} x_{t+1}^i} \right) \# 1 \% CPI \& x,
\]

where \( p_t^i \) is the price of good \( i \) in period \( t \), \( x_t^i \) is the quantity sold of good \( i \) in period \( t \). \( CPI \) is the consumer price index, and \( x \) is the productivity change. The second term in brackets is the weight; it is defined as last period’s revenues from good \( i \) as a fraction of last period’s total revenues. In essence, by considering the effect of price charged today on the index set tomorrow, the firm may be able to earn higher profits tomorrow. For example, by setting a low price for a service today, the weight on the service tomorrow will be reduced when demand is inelastic (as is typical for telecommunication services). The more inelastic demand, the more sensitive are relative revenue shares to small price changes. Even small price changes manipulate the weights sufficiently that profits increase but social welfare declines. He shows that replacing the revenue weights with quantity weights reduces the effect of weight manipulation, especially as demand becomes less elastic. (Further discussion of index issues is contained in Cowan, 1997; Diewert, 1993; and Vickers and Yarrow, 1988a.)

Sappington and Sibley (1992) consider another type of index manipulation; they analyze the effect of price-cap regulation as the FCC planned to implement for AT&T on the firm’s incentive to engage in strategic, non-linear pricing. In their model, the firm sells a single-product and can use two-part tariffs. The price-cap index is represented as

\[
p_t \% \frac{E_t}{Q_{t+1}} \# p_0,
\]

where \( E \) is the fixed part of the two-part tariff, \( p \) is the per unit charge, and \( Q \) is the quantity sold. In other words, average revenue in period \( t \), based on prices charged in \( t \) and on quantities sold in \( t+1 \), must be below the price-cap. They show that incentives are created for the firm to offer a two-part tariff rather than a uniform price, and that such pricing can reduce consumers’ and total surplus. They identify two welfare-improving modifications. First, the index could be specified as a function of the quantities sold in the initial period (replace \( Q_{t+1} \) with \( Q_0 \)). This eliminates the incentive of the firm to reduce the usage charge today (increasing \( Q_{t+1} \)) so that the fixed fee can be increased in the future. The second alternative, which increases welfare even more (and never
reduces consumer surplus) is to require the firm to allow consumers the option to purchase under the original tariff terms before price-cap regulation was implemented.

Armstrong and Vickers (1991) consider another aspect of the role of price structure in a price-cap regulatory scheme. They are interested in whether a multi-product monopolist should be allowed to price discriminate across customer classes, and find that the answer depends on whether a price-cap index is calculated using weights in proportion to demands at uniform prices or whether there is a cap applied to average revenue. Armstrong, Cowan and Vickers (1995) consider how a firm facing a price-cap based on average revenue sets the price structure. (Unlike Sappington and Sibley, 1992, the price-cap is based on prices and quantities this period, so the incentive to set prices this period with an eye to how it would affect the cap next period is eliminated.) They also consider a case where the firm must offer consumers the option of buying at the uniform price. The firm prefers the average revenue constraint to the option constraint, and, of course, likes uniform pricing the least. Consumers overall prefer the option constraint, while ranking the average revenue constraint the least. However, welfare under the various schemes is ambiguous.

The effect of price-cap regulation on quality seems to be a potentially large problem since a decline in quality is in essence a disguised increase in price. (Fraser, 1994, reports that since July 1992, the Australian regulatory body AUSTEL has included quality of service in its evaluation of whether prices have risen in its regulation of Telecom Australia.) Fraser (1994) considers the effect of price-cap regulation on reliability of supply in electric supply; in other words, he incorporates possible changes in the degree of quality when a firm is regulated under a price-cap. He considers two regulatory regimes: in one, the regulator does not incorporate changes in reliability; and in the other, changes in reliability are incorporated into the price-cap. In the latter case, a weighted average of the change in price and the change in reliability must be less than the cap. When reliability is not included in the index, the firm may protect profits by lowering reliability if cost increases must be absorbed (the price-cap is binding). However, when the price-cap is not binding, the firm may increase reliability, given the positive relationship between reliability and price. Including reliability in the price index eliminates the problem of lower reliability when the price-cap is binding, but introduces a new problem of over pricing when the price-cap is not binding. Thus, whichever way the price-cap is formulated, price-cap regulation affects the quality of the service.

Implementation of price-cap regulation has revealed two related, large problems: what is to prevent the regulator from renegotiating the regulatory compact, and how is the end game problem resolved. Part of the reason price-cap regulation works to achieve reductions in cost is by allowing the firm to keep (at least some of) the gains from cost reduction. For this incentive to
work, the regulator must credibly commit not to intervene during the period in which the price-cap is to be in effect. However, such a commitment may not be credible, especially if profits from cost savings rise ‘too much’. If profits rise greatly, the regulatory agency may face pressure from consumer groups to revise the level of the price-cap. In the UK, the electricity distribution regulator decided in June 1995 to revise price-caps, only two months after they had been set, in part because of the gains in productivity (see Studness, 1995) and ‘previously unsuspected financial strength’. Thus, the commitment not to intervene for five years did not even last three months. If regulated firms believe that price-caps will be revised to appropriate the gain from cost savings, or even if there is a chance that revision may occur, then their incentive to invest in cost savings is reduced.

Isaac (1991) considers the problems that may arise when a firm will be regulated under a price-cap for some known, finite period of time, after which rate-of-return regulation will be implemented. He identifies several potential end game problems. First, the firm may manipulate the system by shifting costs into the future. Second, the regulator faces two commitment problems: a commitment not to change the price-cap or intervene in price setting during the price-cap period, and also not use the rate review when moving to rate-of-return regulation to appropriate profits earned by the firm during the price-cap regime. (He discusses these problems in light of the experience of regulation of the Tucson Electric Power Company during the 1980s.)

While several authors have identified these issues as potential problems, not much theoretical work has tackled the issue, with the exception of Armstrong, Rees and Vickers (1995). They derive a model of price-cap regulation with an endogenously determined length of time until the next review. (Most models assume that the price-cap lasts for an exogenously given amount of time and do not consider the effects of a future change in regime on behavior under the current regime.) They identify the following trade off: the longer the amount of time between reviews, the more likely it is the price will be low at the next review, as firms have a higher pay off to investing in cost reduction because they retain profits for a longer period, but the longer is the amount of time during which the price is high. The higher the initial cost when price-cap regulation is implemented, the more likely the former effect is to dominate the latter. They also demonstrate that demand elasticity and the effectiveness of investments in cost reduction are key determinants of the optimal interval between hearings.

Weisman (1993) theoretically analyzes the effect of renegotiation on firm behavior. Given a non zero probability that the regulator will rewrite the regulatory contract and assuming that the probability increases in the regulated firm’s profits, the firm has an incentive to engage in pure waste. In that way the firm may be able to reduce the likelihood of the regulator imposing more stringent regulation. (Sappington, 1980, shows a similar result with respect to
Vogelsang and Finsinger’s 1979 regulatory mechanism.)

Thus, while price-cap regulation does have the benefit of increasing the incentive to firms to invest in cost reduction, it induces a variety of new problems. I now turn to other forms of incentive regulation that may offer improvements to straight price-cap regulation.

14. Alternative Forms of Incentive Regulation

While price-cap regulation is the most commonly discussed and used form of incentive regulation, a few alternatives have been considered as well. The most common alternative is some form of profit-sharing, under which the firm retains some fraction of its profits and rebates the remaining fraction to consumers. Indeed, given the practical problems that arose under price-cap regulation in the UK, profit-sharing regulation was often put forward as a superior regulatory regime that should be adopted. Mayer and Vickers (1996) point out that many of the existing problems would remain under profit-sharing and other, more serious, problems would arise in addition.

An alternative often used in the US is a combination price-cap/rate-of-return/profit-sharing approach to regulation. For example, most US states that have adopted price-cap regulation have included a provision whereby the firm is also subject to a rate-of-return constraint, the form of which Braeutigam and Panzar (1993) term ‘sliding scale’. (In 1992, 22 of 48 states with regional Bell operating companies used a form of sliding scale/price-cap regulation, while 19 states continued with traditional rate-of-return regulation.) Under these provisions, the firm is allowed pricing flexibility below the price-cap, provided that the rate-of-return is not above its cap. (In some states adjustments may be made if the firm’s rate-of-return falls below a floor, as well.) As the rate-of-return rises above the cap, the firm is entitled to a smaller and smaller portion of the increase in profits, typically refunding the rest to consumers. One immediate implication pointed out by Braeutigam and Panzar (1993) is that such an approach eliminates one benefit to price-cap regulation: savings in administrative costs. Under such a hybrid approach, the regulator will still need to calculate rates of return, with all its attendant problems.

Lyon (1996b), using numerical techniques, shows that adding a profit-sharing mechanism to price-caps always increases expected welfare relative to pure price-caps. Welfare may be increased by large amounts of profit-sharing and by allowing firms a greater share of gains than of losses. Profit-sharing is particularly beneficial if the firm’s initial costs are high and cost innovations are difficult to achieve. However, Weisman (1994) shows that the firm prefers profit-sharing to pure price-cap regulation. In essence, this form of regulation gives the regulatory agency an incentive not to take adverse actions against the
firm, since such actions also harm the regulatory agency via the profit-sharing arrangement.

Weisman (1993) considers a ‘modified price-cap’ regulatory regime, which he views as more accurately describing the type of price-cap regulation that has been implemented in the US. Under this regime, the regulator sets a price-cap in the regulated (‘core’) market and allows the firm to retain only part of the profits generated. However, the firm also operates in unregulated (‘non core’) markets as well. To undertake profit-sharing, costs must be allocated across the core and non core markets. Given the difficulty in allocating costs (see the discussion in Section C7), distortions are generated. This type of price-cap/profit-sharing regulation can reduce welfare compared to rate-of-return regulation applied to one market (which also requires that costs be allocated across regulated and unregulated markets). (Braeutigam and Panzar, 1989, early recognized that price-cap regulation can in principle cause a firm to produce efficiently in non core markets and enter these markets only if efficient. However, they analyze a regulatory regime of pure price-cap regulation without profit-sharing, so the problem of allocation of common costs is totally eliminated.)

15. Comparison of Incentive-based Regulation to Other Forms of Price Regulation

Cabral and Riordan (1989) and Clemenz (1991) derive models of price-cap regulation and then proceed to compare the outcomes to several alternative regulatory regimes. Cabral and Riordan (1989) compare cost-based (rate-of-return) regulation with regulatory lag and price-cap regulation. Because there is regulatory lag in their formulation of cost-based regulation, the firm has some incentive to reduce cost. Any increase in profits between the time of cost reduction and the time of the new rate hearing are kept by the firm. However, the longer regulatory lag and the price flexibility increase the incentive to reduce costs relative to the incentive under rate-of-return regulation with regulatory lag. Cabral and Riordan (1989) also compare the level of price under the two alternatives. Under cost-based regulation, at each rate hearing the price is set to cover costs. If the price-cap is initially set below the price that would obtain under cost-based regulation (as it must if the firm is to have an incentive to reduce costs), then for a period prices will be lower under price-cap regulation than under cost-based regulation. However, in future periods the price under price-cap regulation may be higher.

Continuing in the same vein, Pint (1992) considers two differences in price-cap regulation and rate-of-return regulation. She points out that, under price-cap regulation as implemented in the UK, hearings occur at fixed intervals and
use all information available since the previous hearing, while under rate-of-return regulation as implemented in the US, hearings occur only when requested (typically by the firm) and utilize information from a single test year (usually the most recent year for which complete data is available). Assuming that regulation of either form is binding, the firm chooses the level of capital stock, the level of managerial effort (which reduces costs), and, under rate-of-return regulation, when to request a new rate hearing. Both forms of regulation lead to an over investment in capital and underinvestment in effort, especially during periods when cost data are gathered for hearings. Pint (1992) shows that rate-of-return regulation can be improved by using fixed intervals between hearings (continuing to use test year data) if the intervals are not too short, and the welfare gains are primarily realized by the firm. If data on costs since the last hearing are used in conjunction with fixed intervals, welfare increases dramatically, with gains going to consumers (indeed, profits fall). Thus, two aspects of price-cap regulation are seen to increase welfare. (Weisman, 1993, and Baron, 1991, also discuss the degree to which price-cap regulation in practice differs from rate-of-return regulation. Weisman, 1993, points out that in practice (in telecommunications in the US) price-cap regulation involves an element of cost-based regulation. As a result, greater distortions may arise under the hybrid form of regulation than result from pure cost-based regulation.)

Several papers use simulation techniques to compare various regulatory regimes and the degree to which welfare may be increased. Schmalensee (1989a) considers a variety of regulatory regimes, including price-cap regulation, imposed on a risk-neutral, single-product monopolist. He considers what he terms linear regimes, which can be written as

\[ P = \beta + \gamma C + a \]

where \( P \) is the regulated price, \( \beta \) is the base price, \( \gamma \) the cost sharing fraction, \( C \) is observed cost, and \( a \) is the expected unit cost under the pre-reform regulatory regime. That is, any change in cost is passed on linearly to price; the higher is \( \gamma \), the more responsive price is to changes in cost. He finds that price-cap regulation is inferior to cost-based regulation over a wide range of plausible parameter values. While, as many have shown, price-cap regulation provides a higher incentive to invest in cost reduction, if uncertainty is sufficiently high, the price-cap must be set at such a high level to keep the firm profitable that welfare declines. Gasmi, Ivaldi and Laffont (1994) also use simulation techniques to compare welfare under a variety of regulatory regimes, including a price-cap mechanism and price-cap regulation in conjunction with profit-sharing. They characterize their main findings as: (1) pure price-cap regulation leaves substantial rents to the firm; (2) introducing downward price flexibility improves the efficiency of price-cap regulation relative to Schmalensee’s (1989a) best linear regulatory regime; and (3) profit-sharing often yields welfare near the optimal regulated level by partially correcting the distribution distortion of a pure price-cap scheme.
Several works have theoretically examined the performance of price-cap regulation with other types of regulation, and have found it useful to deviate from pure price-cap regulation. For example, performance may be enhanced by offering alternative regulatory mechanisms or alternative pricing schemes, respectively. Lewis and Sappington (1989) suggest that firms should be offered a choice between a variation on price-cap regulation and a profit-sharing form of regulation, with gains going to consumers. Sibley (1989) finds that the regulator uses slugged profits to design its own two-part tariff, and the firm will offer its two-part tariff as well as the regulator’s two-part tariff as an option to consumers. Finally, Schmalensee (1989a) finds that several regulatory mechanisms that are simpler than price-caps can achieve the same gains, particularly if the regulator weights consumer surplus highly.

To recapitulate, theoretical work suggests that some form of incentive regulation can be used to improve the outcome relative to traditional cost-based methods of regulation. However, analysis also shows that, as implemented, incentive regulation often does not replace cost-based methods, and thus new problems arise. Whether incentive-based regulation is superior to cost-based regulation in practice remains to be seen.

16. Empirical Analysis of the Effect of Incentive Regulation

The increasing use of incentive-based regulation has provided economists with a wide range of natural experiments for empirically analyzing the effect of incentive-based regulation on prices, costs and welfare. (Sappington and Weisman, 1996b, concentrating on the telecommunications industry, identify some of the problems that can arise in attempting to do such empirical work, and suggest ways to avoid these problems.) Kridel, Sappington and Weisman (1996) review empirical work on incentive regulation in telecommunications, while Xavier (1995) provides a descriptive narrative of the impact of price-cap regulation in telecommunications, especially in Australia, the UK, and the US. In general, the results have been very encouraging. Productivity, investment in infrastructure, and new service offerings have increased. Prices have generally remained stable or declined slightly, and quality has not declined. However, there is no evidence that incentive-based regulation has led to reduced administrative costs.

Face (1988) examines the use of price-cap regulation for Michigan Bell, estimating that costs savings in 1982 amounted to about $40 million relative to that which would have occurred under rate-of-return regulation. Majumdar (1997) focuses on the impact of incentive regulation on productivity for the case of US local exchange carriers. He finds that pure price-cap regulation leads to
an increase in technical efficiency, but only with a lag; price-cap regulation in conjunction with profit-sharing leads to a smaller but immediate impact on technical efficiency; and finally that a pure profit-sharing scheme actually harms technical efficiency. Mathios and Rogers (1989) find that prices in states that regulated AT&T using a price-cap are 7 to 10 percent lower than in states that use rate-of-return regulation. They use variation in state regulatory mechanisms with a simple price-cap/not price-cap dummy variable. The results are suggestive, but more detailed analysis taking into account the detailed differences in the different state regulatory mechanisms would be desirable.

Despite the variety of problems identified by the theory regarding the implementation of incentive regulation, the empirical studies to date suggest that implementation has led to substantial gains.

E. Regulation of Networks

17. Introduction to the Issues

As competition has been introduced into at least some areas of industries that have long been regulated, a new industry structure is arising. The new structure consists of a situation where a regulated firm controls an essential facility, generally a physical network, to which competitors must have access in order to compete with the regulated firm in at least some markets. For example, in the US much of the competition in local telephone service, at least initially, is expected to come from firms that lease access from the regulated local phone company on which the competitor can send calls. The question is how to appropriately price access to the essential facility. The first-best solution, marginal cost pricing, is generally not feasible because the regulated firm will not recover sufficient revenues to cover the fixed cost of providing the network. (Typically, provision of the network remains a regulated monopoly due to economies of scale.) Thus, the relevant question for economists and the regulator is how to efficiently structure prices subject to the constraint that the regulated firm cover its costs.

Why is regulation necessary in this situation? The economic justification continues to be to reduce the deadweight loss that arises under monopoly. However, the issue is more complicated because the owner of the network competes with others in sectors that are potentially competitive. (When AT&T was broken up, the initial industry structure prohibited the regional operating companies (local service) from competing in these potentially competitive markets (long distance). With the Telecommunications Act of 1996 and deregulation of electricity in the US, the exclusion of network owners from the competitive sectors is declining.) A natural incentive arises for network owners to price access to the network in such a way to give itself an advantage in the
competitive sectors. Another important issue regards the overall regulatory structure and how the competitive market/sector fits in. For example, if prices are high in some potentially competitive sectors in order to subsidize regulated sectors, as was done in the US with regard to long-distance and local service, respectively, then the effect of access/entry on the ability of the firm to maintain such price subsidies must be considered. Thus, the same sustainability issues discussed in Section 10 arise again. I now describe the work that has attempted to characterize the optimal access price.

18. Access Pricing

Baumol and Sidak (1994a) describe the traditional (in the US until the early 1980s) approach for pricing access. Suppose that a regulated firm earns $T$ from a particular service towards its overall revenue shortfall, defined as the difference between total incremental revenue less total incremental cost. Asuming that the price for the final product is fixed, the traditional approach assigns an access fee equal to the average incremental cost of providing access to the competitive firm plus a charge that would leave the regulated firm with its market share weighted share of $T$. That is, if upon granting access a firm is expected to keep two thirds of the market, then the total price for access will be such that the regulated firm earns in total two-thirds of $T$. However, in this situation the firm is not adequately compensated for its common fixed costs. In addition, inefficient entry may occur. The entrant essentially does not have to cover as large a portion of fixed costs as does the regulated firm, and hence can still profitably enter even if its average incremental cost is higher than the regulated firms.

To remedy the entry inefficiency induced by the traditional approach, Baumol, Sidak and Willig have proposed perhaps the most commonly known and used access pricing rule, the efficient component pricing rule (ECPR) or the Baumol-Willig rule. (The rule is generally attributed to Willig, 1979; and Baumol, 1983, and is extensively discussed by Baumol and Sidak, 1994a and 1994c. It has been used in the US by the Interstate Commerce Commission in the rail industry since the early 1980s, by the California Public Utilities Commission in local telephone service since 1989, and in New Zealand in the telecommunications industry since the early 1990s.) Essentially, the rule proposes that access to essential facilities should be priced at the direct cost of providing access plus the opportunity cost to the regulated firm of granting access to a competitor. The opportunity cost to the incumbent is the profit that would have been earned had the incumbent supplied the final product rather than the entering firms. One of the benefits of such a rule is that it leads to efficient entry. In other words, no firm will enter the competitive market unless its marginal cost is less than or equal to the marginal cost of the incumbent regulated firm. Kahn and Taylor (1994) point out an additional potential
benefit. By including the regulated firm’s opportunity cost in the access charge, subsidies mandated by regulators can be maintained. (I call this a potential benefit since many of these cross-subsidies are not efficient. However, this is not always the case, as in Ramsey-Boiteux pricing.) For example, in the days of a common local and long-distance regulated firm, high prices in long-distance service were used to subsidize rates for local service. As discussed in Section 10, entry into the high-profit markets can make the regulatory regime unstable; access pricing under the ECPR can allow entry but sustain the regulatory regime by maintaining the contribution towards the fixed costs provided by that service.

A number of critiques of the ECPR have arisen. In particular, several have questioned the generality of the rule, noting that its efficiency result is contingent on a number of rather strict assumptions. Baumol and Sidak (1994a) and Kahn and Taylor (1994) note that the ECPR gives the efficient solution to pricing access subject to the assumption that the price of the final product is subject to effective regulation or effective competition. Economides and White (1995) develop a rigorous model analyzing the efficiency of the ECPR in the case where the firm that controls the essential facility has market power in the potentially competitive market. They assume that the price for the final product is not regulated, and that the incumbent therefore is able to charge the full monopoly price. Their argument reduces to the point that entry of an inefficient firm, that is a firm with marginal costs above that of the regulated incumbent, can be socially beneficial if entry sufficiently reduces the deadweight loss that arises when the regulated monopolist prices above marginal cost for the final product, while the ECPR in this circumstance acts to protect the monopolist from any competitive challenge. Economides and White (1995) show that under Bertrand or Cournot competition in the final product market, the gain from increased competition can offset some degree of inefficiency on the part of the entrant.

The criticisms of the ECPR highlight an important aspect of the debate: whether the ECPR is efficient depends crucially on the overall regulatory environment, technology, and demand. Is the price for the final product effectively regulated? Laffont and Tirole (1996) emphasize this point: ‘A discussion of an access rule without reference to the rest of the regulatory environment has limited interest. The quality of an access pricing rule depends on the determination of prices for the final products.’ Are products perfect substitutes? Is technology subject to constant returns to scale? It seems clear that the ECPR does not have the wide applicability first claimed. Indeed, many show that while the ECPR arises as a special case of the optimal access rule in certain circumstances, more often than not the optimal access rule may be higher or lower than the EPCR.

Armstrong, Doyle and Vickers (1996) take the ECPR as their framework and analyze how the opportunity cost should be properly measured under
different assumptions regarding demand and supply. They begin with the clear-cut case where the final product is homogeneous; inputs, including access, are used in fixed proportions; no bypass is possible, that is, access is a necessary input into production of the final product and cannot be obtained from any source except the incumbent; and the entrant is a price taker. They then relax some of these assumptions, allowing for product differentiation, variable proportions, and the possibility of bypass. They show that under these new assumptions, the optimal access charge is less than the access charge suggested by the ECPR, because under these assumptions the opportunity cost of access is reduced as the incumbent loses fewer final product sales given entry. However, they also point out that the optimal access charge can be interpreted as an extension of the ECPR, taking into account the ‘displacement ratio’, which is the change in the incumbent’s final product sales over the change in sales of access to competitors as the access price changes slightly. (They also show that, when incorporating the firm’s budget constraint, a positive Ramsey-Boiteux price must be added to the access charge, so that the optimal price subject to the budget constraint is higher than the ECPR access price.) They also show that, when the opportunity cost is analyzed properly, the informational demands of the ECPR are not as low as others have argued; in particular, proper calculation of the opportunity cost requires information regarding demand and supply elasticities.

Armstrong, Doyle and Vickers (1998) consider access pricing in the situation where the price of the final product is not regulated, given that that is the direction in which many industries are headed. They find that, with linear demand and linear competitive supply, and no binding break-even constraint, the optimal price for access is set at marginal cost. More generally the relationship between the optimal access price and marginal cost is ambiguous. In essence the same trade-off identified in Economides and White (1995) is present: setting a high access price raises the price of the final product, resulting in the standard allocative inefficiency associated with monopoly pricing, but it reduces the margin available to competitors, reducing inefficient supply and increasing productive efficiency.

Laffont and Tirole (1994) derive the optimal access price in a wide variety of situations and under pretty general conditions: degrees of effort exerted by the incumbent in operating the network, private information on the part of the incumbent regarding technology, various informational regimes, the presence and absence of government transfers, various degrees of market power on the part of the entrant, various regulatory regimes, and under the possibility of bypass. They also compare the optimal access price to that suggested by the ECPR. They identify the following assumptions as necessary to ensure the efficiency of the ECPR: (1) the regulated firm’s price for the final product is based on marginal-cost pricing; (2) the products produced by the regulated firm and the entrant are perfect substitutes; (3) production of the final product is
characterized by constant returns to scale; (4) the entrant has no market power; and (5) the regulated firm’s marginal cost can be accurately observed. If any of these fail to hold, then the access price must be adjusted from the ECPR level to achieve efficiency.

Larson and Lehman (1997) demonstrate that ECPR pricing can be derived as a special case of Ramsey-Boiteux pricing with interdependent demands, but also show that ECPR is not always efficient. In particular, in order for a Ramsey-Boiteux pricing structure to arise, the following assumptions must hold: (1) entrants must have a fixed proportions technology; (2) the essential facility must be strict, that is have a zero elasticity of substitutions for other inputs; (3) perfect competition must exist in the non essential inputs; (4) the regulated firm and entrants must face the same elasticity of demand for the final product; (5) the entrants set price at the level charged by the regulated firm; (6) Bertrand competition ensues in the downstream market, which in combination with assumption (5) results in equal market shares for the regulated firm and entrants; and (7) there is symmetry of weighted income effects. They then discuss the many reasons why these assumptions are likely to be violated.

Finally, Laffont and Tirole (1996) consider a completely different approach to regulating interconnection. They suggest that access be regulated under a global price-cap. Under this approach, the price-cap would apply the weighted sum of price changes for final products and price changes for intermediate products sold to rivals, including access. They argue that one benefit is that it requires no more informational requirements than existing regulatory policies. The firm will adopt the optimal Ramsey-Boiteux price structure given its information about demand and cost; the regulator does not require information on marginal costs nor demand elasticities. (Ramsey-Boiteux prices obtain so long as all goods, including access, are included in the definition of the price cap and the weights are exogenously set at the level of output to be realized.) They also consider a global price-cap that also imposes the ECPR as a price ceiling on the access price, though the informational requirements of the ECPR are severe under more realistic assumptions, as shown by Armstrong, Doyle and Vickers (1998). The benefits identified by Laffont and Tirole (1996) from adding an ECPR price ceiling is to provide a better setting of the weights in the price-cap and to limit predatory behavior (the incumbent may increase the access price and reduce the final output price, squeezing out entrants) by tying the access and output prices.

The theoretical work suggests that a general approach to pricing access at the direct marginal cost of access plus the opportunity cost is appropriate. However, it appears that calculating the opportunity cost is not as easy as originally proposed. In addition, the work illustrates the importance of taking into account the entire regulatory regime, including the regime covering the competitive sector. (I found no empirical work regarding access prices.)
F. Conclusion

19. Regulation in Practice

As regulation theory has advanced, especially in the last decade or two, regulation in practice has changed tremendously. The US saw immense deregulation occur beginning in the 1970s, mostly in industries that were never believed to be natural monopolies but had been regulated for other reasons. (For example, regulation of trucking in the US largely developed in order to protect regulation of railroads, which were thought to be natural monopolies.) In particular, deregulation occurred in railroad and trucking, air passenger transportation, natural gas, electricity, telecommunications, and cable television. (About the only move in the US in the opposite direction was the 1992 Cable Consumer Protection and Competition Act, which re regulated cable television in response to the increase in cable prices after deregulation in the Cable Communications Policy Act of 1984. It is not clear how regulation affected the quality adjusted price; Hazlett, 1993, for example, believes the quality-adjusted price declined under deregulation, while Otsuka, 1997, suggests that the quality-adjusted price was kept low under regulation. However, cable was rederegulated with the passage of the Telecommunications Act of 1996.) Other countries have seen an impressive move away from handling natural monopolies by nationalizing them to regulation of privatized firms. These countries, especially the United Kingdom, have led the way towards the use of incentive regulation.

Not only has the US moved towards deregulation, but the form of regulation has been changing rapidly over the 1980s and 1990s. There is a general movement away from cost-based regulation towards incentive-based regulation, in an effort to achieve a better outcome. The United Kingdom led the way with adoption of price-cap regulation in 1984, after privatizing its telecommunications industry. Privatization of other industries spread the use of price-cap regulation to electricity, gas and water. Price-cap regulation in practice in the United Kingdom has included a bit of a mix of cost-based regulation. For example, gas, water and electricity utilities can pass on increases in costs of inputs outside their control. Nonetheless, this method of regulation has proved to be revolutionary, leading to changes in the approach to regulation in many countries.
Acknowledgments

I gratefully acknowledge the comments of Ken Boyer and of an anonymous referee. Their comments significantly improved the content and organization of the paper and extended the bibliography. Chris Haan, Vivian Lei and Aaron White provided valuable assistance in compiling the bibliography.

Bibliography on Price Regulation: A (Non-Technical) Overview (5200)


Bourgoignie, Thierry (1973), ‘Overheid en Bedrijfsleven/La Reglementation des Prix en Belgique (Price Regulation in Belgium)’, 27 Economisch en Sociaal Tijdschrift, 385-400.


Brennan, Michael J. and Schwartz, Eduardo S. (1982), ‘Consistent Regulatory Policy under
Brennan, Timothy J. (1991a), ‘Regulating by Capping Prices’, 1 Journal of Regulatory Economics,
133-148.
Broadman, Harry G. and Kalt, Joseph P. (1989), ‘How Natural is Monopoly? The Case of Bypass in
Broadman, Harry G. and Montgomery, David W. (1983), Natural Gas Markets after Deregulation,
Washington, Resources for the Future.
up Bell: Essays on Industrial Organization and Regulation, Amsterdam, North-Holland.
Monopoly: Bertrand Revisited by Cournot’, in Evans, David S. (ed.), Breaking up Bell: Essays
on Industrial Organization and Regulation, Amsterdam, North-Holland.
Brown, Allan and Cave, Martin (1992), ‘The Economics of Television Regulation: A Survey with
Application to Australia’, 68 Economic Record, 377-394.
Equilibrium Analysis’, in Aliprantis, C.D., Burkinaf, O. and Rothman (eds), Advances in
Equilibrium Theory: Proceedings of the Conference on General Equilibrium Theory, New York,
Springer, 43-54.
York, St. Martin’s Press, 77-83.
Brown, Lorenzo (1988), ‘Modelling Diversified Utilities and Cross Subsidization’, 10 Resources and
Energy, 213-224.
Cambridge University Press.
Burness, H. Stuart and Patrick, Robert H. (1991), ‘Peak-Load Pricing with Continuous and
Burness, H. Stuart, Montgomery, David W. and Quirk, James P. (1980), ‘Capital Contracting and the
Caillaud, Bernard, Guesnerie, Roger, Rey, Patrick and Tirole, Jean (1988), ‘Government Intervention


Clark, John Maurice (1923), Studies in the Economics of Overhead Costs, Chicago, University of Chicago Press.

MA, and Dordrecht Kluwer Academic, 57-79.


Economides, Nicholas and White, Larry (1998), The Inefficiency of the ECPR yet Again: A Reply to Larson, 43 Antitrust Bulletin.
Einhorn, Michael A. (1991b), ‘Price-Caps and Incentive Regulation in Telecommunications:


Hillman, Jordan Jay and Braeutigam, Ronald R. (1989), Price Level Regulation for Diversified Public
Utilities, Boston, Kluwer.


Labus Miriljub (1990), 'Devalvacija i Administrativna Kontrola Cena (Devaluation and Administrative Control of Prices)', 1 Ekonomiska Misao, 1-24.


MacKie-Mason, Jeffrey K. and Waterman, David (eds) (1998), Telephony, the Internet, and the


Price Regulation: A (Non-Technical) Overview


Peltzman, Sam (1976), ‘Toward a More General Theory of Regulation’, 19 Journal of Law and
Price Regulation: A (Non-Technical) Overview

5200


Price Regulation: A (Non-Technical) Overview


Spiller, Pablo T. and Vogelsang, Ingo (1993), ‘Regulation Without Commitment: Price Regulation of
Regulation’, 19 Telecommunications Policy, 151-160.
Stigler, George J. (1975), The Citizen and the State. Essays on Regulation, Chicago, University of
Stigler, George J. and Friedland, Claire (1962), ‘What Can Regulators Regulate? The Case of
Stiglitz, Joseph E. (1977), ‘Monopoly, Nonlinear Pricing and Imperfect Information’, 44 Review of
Economic Studies, 407-430.
Stolley, Kenneth R. (1981), ‘Price Controls on Nonrenewable Resources when Capacity is
Stout, Lynn A. and Barnes, David D. (1992), Economic Foundations of Regulation and Antitrust
Law, St Paul, West Publishing.
41 Southern Economic Journal, 149-151.
Forthnightly, 36-37.
31 Comparative Economic Studies, 10-52.
Sweeney, George (1982), ‘Welfare Implications of Fully Distributed Cost Pricing Applied to Partially
International Journal of Industrial Organization, 247-261.
Taussig, Frank W. (1913), ‘Railway Rates and Joint Costs’, 27 Quarterly Journal of Economics,
692-694.
Tawada, Makoto and Katayama, Sei Ichi (1990), ‘On the Technical Efficiency under Regulation: A
11 Review of Industrial Organization, 459-471.
Thomadakis, Stavros B. (1982), ‘Price Regulation under Uncertainty in an Asymmetric Decision
Train, Kenneth E. (1991), Optimal Regulation: The Economic Theory of Natural Monopoly,
Cambridge, MA, MIT Press.
Turvey, Ralph (1968), Optimal Pricing and Investment in Electricity Supply, Cambridge, MA, MIT
Press.
Regulation, 203-224.
Weiss, Lawrence and Klass, Michael W. (1986), Regulatory Reform: WhatActually Happened,
Boston, Little Brown.


Absorption’, 24 Regional Science and Urban Economics, 185-205.
Zajac, Edward E. (1972), ‘Note on “Gold-Plating” or “Rate Base Padding”’, 1 Bell Journal of Economics.