2 Price regulation: theory and performance

Janet S. Netz

1. OVERVIEW

1.1 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.\(^1\) 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 survey is organized as follows. In Section 1.2, 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. Those familiar with the theory of natural monopolies can skip this section. Perhaps the most widely used early form of price regulation was rate-of-return regulation, which I describe in Section 2.1. 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 2.2, Averch and Johnson (1962) show that such an approach to regulation encourages the public utility to overinvest in capital and to expand into other markets in order to increase the capital in its “rate base”, hence increasing the aggregate amount of profits it

---

\(^1\) 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 “Economic Theories of Regulation”, chapter 1 in this volume, for a discussion of the political and economic forces that have led to regulation.
could earn subject to the regulatory constraint. In Section 2.3, I describe extensions to the Averch-Johnson model that make the model more descriptive of actual regulation. I discuss the potential benefits of rate-of-return regulation versus the market outcome in Section 2.4, and summarize empirical work on the existence and magnitude of the negative effects induced by rate-of-return regulation in Section 2.5. I summarize in Section 2.6.

Regulation theory has more recently focused on designing regulatory mechanisms that do not provide the public utility with the adverse incentives described above. In Section 3, I consider alternative forms of price regulation that fall under the rubric “incentive regulation”, focusing on price-cap regulation in Sections 3.2 and 3.3 and on other forms of incentive regulation such as profit-sharing in Section 3.4. In Section 3.5, I highlight some of the attempts to compare incentive regulation with other forms of regulation, in terms of both their welfare effects and the feasibility of implementation. Section 3.6 discusses empirical attempts to measure the impact of incentive regulation.

In Section 4, I discuss regulation of networks, where the focus is on access pricing in situations where an incumbent firm provides access to an essential facility, generally a physical network, to rivals, as discussed in Section 4.2, and in situations where rival firms want access to each other’s networks due to network externalities, as discussed in Section 4.3. In Section 5, I discuss the overlap and boundaries between price regulation and competition (antitrust) policy. A firm that controls a network that is an input to provision of a downstream service can use its network market power to obtain market power in the downstream market. While this is typically seen as an issue for competition (antitrust) policy, the presence of regulation of the network changes the incentives of the firm. I conclude in Section 6.

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 volume of the Encyclopedia of Law and Economics cover these related topics, I omit discussion of: the theory of regulation, or why different forms of price regulation might be implemented (see “Economic Theories of Regulation”), as well as regulation in particular sectors (see “Telecommunications Regulation”, “Electricity Regulation”, “Regulation of the Natural Gas Industry”, “Regulation of Water and Wastewater”, and “Regulation of the Global Transport Industry: An Institutional Account”). I attempt to balance discussion of the theoretical work with empirical analysis of the effect of various types of price regulation mechanisms.

1.2 The Theory of How Regulation Can Improve Social Welfare in the Presence of a Natural Monopoly

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. Typically this is the case when there is a significant fixed cost of providing the service; in regulated industries, this fixed cost is often the cost of a physical network, e.g., electric, phone, or cable lines running to individual houses and businesses. When the fixed cost is significant relative to demand, it is cheaper for a single firm to serve the market so as not to duplicate the fixed costs, e.g., not to have two (or more) sets of phone (or electric or cable) lines running to each house and business.

I illustrate a natural monopoly arising from economies of scale over the relevant range of production for a firm that sells a single product in Figure 2.1. The average total cost curve is shown to be everywhere declining (and hence the marginal cost curve is beneath the average total cost curve), reflecting economies of scale; with the technology reflected in the cost curves, cost declines as output increases. Thus, the industry is a natural monopoly; any market structure involving several firms would involve the unnecessary duplication of fixed costs.

Figure 2.1 Market outcome versus regulated outcomes

for one firm to serve the market than for several firms to serve the market. Typically this is the case when there is a significant fixed cost of providing the service; in regulated industries, this fixed cost is often the cost of a physical network, e.g., electric, phone, or cable lines running to individual houses and businesses. When the fixed cost is significant relative to demand, it is cheaper for a single firm to serve the market so as not to duplicate the fixed costs, e.g., not to have two (or more) sets of phone (or electric or cable) lines running to each house and business.

I illustrate a natural monopoly arising from economies of scale over the relevant range of production for a firm that sells a single product in Figure 2.1. The average total cost curve is shown to be everywhere declining (and hence the marginal cost curve is beneath the average total cost curve), reflecting economies of scale; with the technology reflected in the cost curves, cost declines as output increases. Thus, the industry is a natural monopoly; any market structure involving several firms would involve the unnecessary duplication of fixed costs.

See Panzar (1989) for a complete, technical discussion of the technology that leads to a natural monopoly. The conditions, especially for a multi-product firm, are considerably more complicated than the simple graphical analysis presented in Figure 2.1.
Assume that the firm (and regulator) can charge only a single price; that is, price discrimination (charging different prices to different consumers) is not allowed nor is it feasible. 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 $Q_M$, where the $M$ subscript denotes “Market”. Here the firm is making profits, as its average total cost of producing $Q_M$ (indicated in the figure by the horizontal line labeled $ATC_M$) is less than the price charged ($P_M$). This outcome achieves neither allocative efficiency nor productive efficiency.

An allocatively efficient outcome occurs where the cost of producing one more unit of the product or service is equal to the benefit of consuming one more unit of the product or service. At this point, social welfare is maximized. This is indicated in Figure 2.1 where the marginal cost curve intersects the demand curve, and corresponds to a price of $P_E$ and a quantity of $Q_E$, where the $E$ subscript denotes “Efficient”. Productive efficiency occurs when the product or service is produced at the lowest possible cost, which would be indicated by producing at the minimum of the average total cost curve. Typically, in the presence of a natural monopoly, the issue is that demand is not sufficient to reach the point of minimum cost.\(^3\)

Price regulation can theoretically lead to the social optimum if regulators specify that price be set equal to $P_E$, at which point allocative efficiency is achieved.\(^4\) At the efficient outcome, social welfare has increased by the amount of the shaded triangle relative to the market outcome. However, a firm that charges $P_E$ and produces at $Q_E$ will not generate sufficient revenue to cover the costs of production; this can be seen graphically in that the average total cost for producing $Q_E$ units is above $P_E$. 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.\(^5,6\)

---

\(^3\) Often, the term productive efficiency is used more generally to indicate that the firm is producing a given outcome at the lowest possible cost. That is, while it may be that the average total cost is lower if output is increased, output may be described as productively efficient if that amount cannot be produced more cheaply. This type of efficiency is more precisely called technical efficiency.

\(^4\) In addition, the outcome has moved towards productive efficiency. Pure productive efficiency cannot be achieved because demand is not of sufficient magnitude for production to occur at the minimum of the average total cost curve.

\(^5\) 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.

\(^6\) Pricing mechanisms other than a single price – e.g., third-degree price
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 the shaded area less the triangle outlined in bold, relative to the market outcome. Society is still losing the area outlined in bold, but this may be acceptable relative to the political cost of providing the firm with a subsidy equal to the firm’s fixed costs.

Thus, in theory, price regulation 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 intersect, though in practice it may only be possible to move towards the social optimum, not to achieve it.

2. THE IMPACT OF RATE-OF-RETURN REGULATION ON INCENTIVES

2.1 Introduction to Rate-of-Return Regulation

Regulation of prices, as historically practiced in the US and a number of other countries, has often involved rate-of-return regulation. Under such regulation, the regulatory agency sets prices in such a way that the public utility earns the allowed (“fair”) rate of return on its investment. The allowed rate of return can be set in such a way that it induces the outcome $P_F$, $Q_F$ as indicated in Figure 2.1. In this case, social welfare is increased substantially relative to the market outcome (the increase in social welfare is graphically indicated by the shaded triangle less the boldly outlined triangle).

This form of regulation has been quite prevalent in the US, and while not as widely used today, is still common. For example, virtually the entire electricity industry has been subject to rate-of-return regulation, and as of 2005, over half the states continued to use this form of regulation (Vitaliano and Stella 2009 and Casten 2007). The US telecommunications industry relied on rate-of-return regulation until the break-up of AT&T, and about half the states continued to use rate-of-return regulation as of 2004 (Knittel 2004). This type of regulation was used in Spain for the electric industry until 1988. Japan has traditionally used rate-of-return regulation for its electric industry.\(^7\)

\(^7\) In 1996, Japan refined its regulation to take into account the performance of firms relative to other utilities in an effort to increase efficiency.
While this form of regulation is seemingly simple, and seems as if it would achieve feasible average cost prices, in practice the outcome is often less than ideal. The problem is generally caused because the regulation changes the incentives faced by the firm.

### 2.2 The Over-investment Incentive

Perhaps the earliest reaction to the change in incentives was illustrated by Averch and Johnson (1962). They model the regulated firm as maximizing 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 (and 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 will overinvest in capital. The firm will do this in two ways if feasible: (1) the firm will use too much capital relative to labor and other inputs, and thus will not minimize costs, and (2) the firm will expand into other markets, even if it operates at a loss in these markets, and hence may drive competitive firms out of these markets (or discourage them from entering), if at least some of the capital used in the other markets is included in its rate base. These incentives arise because the firm doesn’t earn a profit from selling its product or service, but rather makes a profit for each additional unit of capital used.

Assume that a firm makes a single product with inputs capital ($k$) and labor ($l$). Assume as well that the two inputs are sold in competitive markets at prices $r$ and $w$, respectively. Finally, assume that the firm earns revenues from its use of inputs according to the function $R(k,l)$. Mathematically, the regulated firm maximizes its profits ($R(k,l) - wl - rk$) subject to the constraint that its earnings are less than the allowed rate of return

$$\left( \frac{R(k,l) - wl}{k} \right) \leq s,$$

where $s$ is the allowed rate of return. Then the firm’s constrained maximization problem is given by the Lagrangian

---

8 The rate base is the amount of capital on which the firm can earn the allowed rate of return.

9 Baron (1989) points out that the regulated firm would like the price of the product or service 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.

10 The implications are unchanged if the firm makes multiple products with more than two inputs.
where $\lambda$ is the Lagrange multiplier. The first-order conditions can be rewritten to give

$$
\frac{MP_k}{MP_l} = \frac{r}{w} - \frac{\lambda}{1-\lambda} \left( \frac{s-r}{w} \right) < \frac{r}{w},
$$

where $MP_i$ is the marginal product of input $i$ (the extra output produced by employing one more unit of input $i$).\(^{11}\) For a regulated firm, the ratio of the marginal product of capital for labor is less than the ratio of the marginal cost of capital for labor. For an unregulated firm, however, the constraint is not present, and the second term in the middle expression does not occur. Instead, an unregulated firm uses capital and labor such that the ratio of the marginal product of capital to labor is equal to the ratio of the marginal cost of capital to labor.

Consider, first, the meaning of the first-order conditions for an unregulated firm. This firm will combine labor and capital in such a way that 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, earns revenues not only by selling the output produced by the inputs, but also on the input 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, thus producing at higher than minimum cost. The degree to which a firm overcapitalizes depends, inter alia, on its ability to substitute between inputs. The more substitution the technology allows, the more inefficient a firm will be. Takayama (1969) shows that as the allowed rate of return is reduced, the firm increases its use of capital relative to the efficient amount.

Now consider a firm that may operate in several markets, not all of which are regulated, and 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 costs incurred by the firm as a whole that are not directly attributable to particular lines of business). 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

\[11\] The last inequality follows if $\lambda \in (0,1)$. See Averch and Johnson (1962) and Baumol and Klevorick (1970). See also Takayama (1969), who argues that $\lambda$ need not be less than 1, but that the over-capitalization result still holds.
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, unregulated markets. Suppose that the regulated firm is not competitive in the unregulated market: that is, its costs are above rival firms’ costs. Then the regulated firm will lose money on each unit that it sells in the unregulated market. However, for each unit of capital that the regulated firm uses to produce the unregulated product, it earns revenue because it loosens the rate-of-return constraint. Then the firm will operate in the unregulated market as long as the profits from using capital in the unregulated market (which is equal to the portion of the capital in the other market that is allowed into the rate base, times the allowed rate of return less the cost of capital) are less than the losses from operating in the market. 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 discouraged from entering markets in which they have lower costs.

In theory, the regulator can set the allowed rate of return to maximize social welfare taking into account the Averch-Johnson incentive to overcapitalize. Callen, Mathewson, and Mohring (1976) consider this question, inter alia. However, they point out that this question in some ways misses the boat. If the regulator had sufficient information to determine this socially optimal allowed rate of return, the regulator could dictate the efficient level of inputs and avoid the overcapitalization effect to begin with.

### 2.3 Regulatory Lag and Cost Disallowances

The Averch-Johnson effect is developed in a model of active, concurrent regulation. However, in practice rate-of-return regulation is largely implemented prospectively. That is, expected demand and costs are forecast, and prices are set for the coming period. Rate hearings occur periodically, and are typically based on historic costs. Between the time when the regulated price is set and the rate hearing, a firm can profit not only on its use of capital but also from any cost reductions it can achieve. The latter re-introduces an incentive for the firm to minimize cost, including by utilizing the efficient ratio of capital to labor.


---

12 Averch and Johnson (1962) argue that the descriptive evidence suggests that AT&T behaved along the lines predicted by their model in terms of overinvesting in capital and moving into other markets.
lag since, 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 re-introduces to the firm the incentive to reduce costs, though the incentive is not as large as it would be if the firm appropriated the gains from cost reductions forever. Thus, with regulatory lag, the Averch-Johnson effect is mitigated, although not eliminated.

Regulators have another tool at their disposal to counteract regulated firms’ incentive to overcapitalize: the regulator can disallow capital expenditures. Encinosa and Sappington (1995) and Lyon (1991, 1992) examine a firm’s incentive to over- or underinvest in capital in 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 about 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 find that rewards as well as penalties are optimal. Lyon (1992) examines prudence reviews in the context of contracts for variable inputs, and finds that prudence 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.

2.4 The Benefits of Rate-of-Return Regulation Relative to the Market Outcome

Callen, Mathewson, and Mohring (1976) consider the benefit that derives from rate-of-return regulation for various parameterizations of a constant elasticity demand equation and a Cobb-Douglas production function, though they do not incorporate an estimate of the costs of regulation. While they find

13 A Cobb-Douglas production function has the form \( Y = AL^\alpha K^\beta \), where \( Y \) is output, \( A \) is a constant, \( L \) is the amount of labor, and \( K \) is the amount of capital. The percentage change in output given a 1% increase in labor (capital) is given by \( \alpha (\beta) \). This flexible functional form can encompass constant returns to scale (if \( \alpha + \beta = 1 \)), increasing returns to scale (if \( \alpha + \beta > 1 \)), or decreasing returns to scale (if \( \alpha + \beta < 1 \)). It can be extended to account for more than two inputs.
that the degree of Averch-Johnson cost inefficiencies are significant, the benefits that arise from rate-of-return regulation are nonetheless substantial, ranging from a low of 11% to a high of 221% when the allowed rate of return is set taking into account the Averch-Johnson effect. They also evaluate the welfare benefits from rate-of-return regulation when the allowed rate of return is not set optimally (recognizing the seemingly prohibitive informational requirements of determining the optimal level), and find benefits ranging from 8% to 217%. Thus, while the relatively simple rate-of-return approach to regulation induces inefficiencies, the benefits from regulation, versus an unconstrained market outcome, are substantial.

2.5 Empirical Tests Related to Rate-of-Return Regulation

Empirical work testing the Averch-Johnson overcapitalization result has focused on electricity generation. While the results of these tests have been mixed, there is more evidence in support of than against the presence of an Averch-Johnson effect.

In the US, Courville (1974), Hayashi and Trapani (1976), and Spann (1974) all find support for the Averch-Johnson effect. Courville’s estimate is that actual costs are over 11% higher than minimum costs. Hayashi and Trapani (1976) also find that tightening regulation increases the distortion, which is consistent with Takayama (1969). Boyes (1976), on the other hand, does not find support for the Averch-Johnson effect. His approach amounts to estimating the value of the Lagrange multiplier; he cannot reject the null hypothesis that it is zero. While he interprets this as evidence against the finding of an Averch-Johnson effect, the result is consistent with rate-of-return regulation that sets the allowed rate of return higher than what the public utility could make if it were not subject to regulation. 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. Finally, Hsu and Chen (1990) find empirical evidence for the presence of an Averch-Johnson effect for the Taiwan Power Company.

Oum and Zhang (1995), the only study I found outside the electric industry, 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.

More recently, several authors have attempted to test for (in)efficiency more generally. These papers are testing (a) whether the optimal quantity is
being produced and (b) whether output is being produced at minimum cost. Atkinson and Halvorsen (1980, hereafter A&H) adjust their approach to take into account that rate-of-return regulation is not continuously enforced and to incorporate fuel adjustment clauses. The fuel adjustment clauses cause an overuse of fuel relative to labor and to capital, as standard rate-of-return regulation causes an overuse of capital relative to labor (and to fuel). The net effect of the Averch-Johnson effect and the fuel adjustment clause on the relative use of fuel and capital is theoretically indeterminate. A&H examine electric plants in the US from 1973. They find support for the Averch-Johnson effect: capital is used too intensively relative to labor. They also find that fuel adjustment clauses lead to an overuse of fuel relative to labor. The two effects offset each other, and they cannot reject the null hypothesis that capital and fuel are used in efficient proportions.

Vitaliano and Stella (2009, hereafter V&S) examine 337 US electric generating plants from 1970, while Rungsuriyawiboon and Stefanou (2007, hereafter R&S) examine 72 US electric generating plants from 1986–99. These firms do not assume technical efficiency (that is, that the firm cannot produce more output with the inputs it is employing), which theoretical and much empirical literature do. Both find strong evidence of both kinds of inefficiency: the plants are neither technically nor allocatively efficient. The Averch-Johnson effect is one of allocative inefficiency, where the firm uses too much capital to produce the given amount of output. R&S also find, in contrast to A&H, that plants use too little fuel relative to labor. This may reflect the substantially different time periods covered by the two studies (1986–99 versus 1970).

Unlike other studies, Vitaliano and Stella try to relate the degree of overcapitalization to what they call the capital subsidy, or the difference between the allowed rate of return and the cost of capital. The capital subsidy provides virtually no explanatory power. Based on the lack of explanatory power, V&S conclude that there is no support for the Averch-Johnson effect, despite their finding of overcapitalization. The lack of explanatory power, however, may be due to data issues. It is not clear that there is much variation in the degree of overcapitalization; unfortunately, V&S report only the mean value of overcapitalization and not the variance. In addition, the measure of the capital subsidy is rather crude. No attempt is made to adjust the accounting data to effectively measure the economic rate of return.

14 R&S investigate the degree of inefficiency across states that are actively implementing deregulation plans and those that are not. Generally speaking, states that are deregulating have electric plants that are less technically inefficient than states that are not deregulating, but the degree of allocative inefficiency (the Averch-Johnson effect of over-capitalization) is not much different across the two groups of plants.
Finally, V&S offer no alternative explanation for the observed allocative inefficiency of the electricity firms.

Given the theory and evidence regarding the disincentives to minimize costs, traditional rate-of-return regulation has been altered in a variety of ways to incorporate incentives for cost minimization. Knittel (2002) investigates the impact of several alternative types of regulation on efficiency of electric generation in the US. He examines the effect of two programs that directly reward efficiency, heat-rate programs and equivalent availability factor programs. The former set prices or profits conditional on the firm's average heat rate, and the latter tie prices or profits to the amount of time that plants are online. This can reduce generation costs if low-cost plants are available for generation. Another form of regulation that had been adopted to a small extent in the electric industry at the time of his study allowed firms to appropriate some of the cost savings they can generate; this form of regulation was called a rate-of-return range alternative. Finally, some states adopted partial fuel pass-through programs, under which firms could pass on a fraction of fuel cost increases, as well as profiting, to some degree, from declines in fuel costs.

Knittel examines US electricity-generating plants from 1981 to 1996. For both coal and gas plants, he finds that the modal plant operates efficiently, but the mean plant is inefficient. The mean coal- and gas-powered plant, not operating under any of the alternative modes of regulation described in the previous paragraph, produces 17.4% and 5.4% less, respectively, than it could given its inputs. He finds that the three alternative forms of regulation described in the previous paragraph – equivalent availability factor program, heat rate program, and rate-of-return range – increase efficiency. These forms of regulation increase output by 10.5%, 9.5%, and 4.0%, respectively, for coal-powered plants. For gas-powered plants, those operating under equivalent availability factor and heat-rate programs (no gas-powered plants operate under one or the other) are slightly more efficient, increasing output by 1.9%. For both types of plants, partial fuel pass-through programs lead to more efficiency than complete fuel pass-through programs. Knittel’s work clearly indicates that modifying traditional rate-of-return regulation can mitigate the disincentive to minimize costs, leading to more efficiency than under traditional rate-of-return regulation.

2.6 Summary

Overall, theoretical models support the Averch-Johnson findings of a tendency towards overinvestment in capital by firms subject to a rate-of-return

15 The more efficiently a plant generates electricity, the lower is its heat rate, which is essentially the amount of energy wasted in the production process.
constraint, though introducing some modifications of the framework, such as regulatory lag and cost disallowances, reduces the overinvestment effect to some extent. It appears that rate-of-return regulation can improve social welfare even in the presence of the Averch-Johnson effect. Finally, there is more empirical support for the Averch-Johnson effect than against it.

3. INCENTIVE REGULATION

3.1 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.\textsuperscript{16} Under rate-of-return regulation, 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 needn’t worry about rising costs, because those costs will be covered by a higher regulated price. Part of the motivation behind incentive regulation (sometimes referred to as performance-based regulation) is to provide the firm with the motivation to behave in a manner more consistent with the social optimum, which means minimizing costs. See Crew and Kleindorfer (1996), Hemphill, Meitzen, and Schoech (2003), and Joskow (2008) for discussions of the use of incentive regulation. There are many variations of incentive-based regulation, but the most commonly used include price-caps and earnings sharing, which I discuss in more detail in Sections 3.2, 3.3, and 3.4 below.

Price-cap regulation was designed to eliminate the cost disincentives other forms of price regulation caused; in addition, price-cap regulation reduces regulatory administrative costs (at least in theory). Under price-cost regulation, the regulator sets a price ceiling, and firms can charge prices up to the ceiling. The firm retains at least a portion of any cost savings, reinstating the incentive of the firm to minimize cost. In addition, in a pure price-cap regime, the regulated market is de-coupled from other markets, eliminating the inefficient incentives firms subject to rate-of-return regulation face with regard to entering markets in which they are not competitive (see the discussion in Section 2.2). Finally, price-cap regulation may have added benefits in industries where technology is changing rapidly, since the price ceiling is often set as a formula requiring the firm to reduce prices each year. Maintaining rate-of-return regulation when

---

\textsuperscript{16} As discussed in Section 2.3, 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, it is not eliminated.
technology is changing rapidly 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.\textsuperscript{17}

Price-cap regulation was adopted in the UK after 1984 to regulate natural monopoly industries that were being privatized; the UK now applies this form of regulation to electricity, natural gas, telecommunications, and water.\textsuperscript{18} According to Hemphill, Meitzen, and Schoech (2003), the first use of incentive regulation in the US was adopted to regulate railways in the 1980s. Price-cap regulation has been adopted for use in the US, primarily in telecommunications after the break-up of AT&T, at the federal level and by about half the states by 2004. By 1993, 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 3.4. The US Federal Communications Commission 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. The electricity and natural gas industries in the US have made some moves from rate-of-return regulation to incentive regulations. Spain transitioned its electric industry from cost-based to incentive-based regulation in 1988, and Canada is moving to price-cap mechanisms for regulation of natural gas. Many other countries have also adopted incentive-based regulation, including Australia, Denmark, France, Germany, Hong Kong, Mexico, the Netherlands, and Sweden.

### 3.2 Price-Cap Regulation

To give an idea of how price-cap regulation works in general, consider the following scenario. The regulatory agency typically chooses two values in setting the price-cap – the initial price-cap and an adjustment term. The adjustment term generally reflects technological progress that brings about reductions in costs. Subsequent price-caps are equal to the initial price-cap less the adjustment term. During the time the price-cap is in effect, the regulated firm may either choose a price below or equal to the price-cap.\textsuperscript{19} The firm can

\textsuperscript{17} This benefit of price-cap regulation over rate-of-return regulation in the face of rapid technological progress may explain why telecommunications regulation adopted price-cap regulation more quickly than other industries in the US. Einhorn (1991), Mitchell and Vogelsang (1997), and Sappington and Weisman (1996a), provide comprehensive treatment of incentive regulation in the telecommunications industry.

\textsuperscript{18} Armstrong, Cowan, and Vickers (1994) and Beesley and Littlechild (1989) describe incentive regulation in the UK.

\textsuperscript{19} Note that the firm is granted a high degree of pricing flexibility; in particular,
request a rate hearing, which it might do, for example, if its costs were higher than those contemplated when the price-cap was set. 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.

Cabral and Riordan (1989) set out a fairly simple model of 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, as long as higher effort increases the likelihood that marginal cost will fall. Let $C$ denote the firm’s original marginal cost; let $(P_0 - x)$ denote the price-cap, which is the initial price-cap (denoted $P_0$) less the adjustment factor ($x$) and is set below $C$; and $M(C)$ denote the unconstrained monopoly price for a marginal cost of $C$. Then the price charged by the regulated firm is a function of the price-cap and of marginal cost. If costs are above the price-cap, the firm asks 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, then the firm prices at the monopoly price.

Cabral and Riordan (1989) derive several illustrative results. First, in the case where the firm knows by how much marginal cost will be reduced for a given investment in cost-reducing effort, they show that if the price-cap is too low, then the firm has no incentive to invest in cost reduction because the return 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 a tighter price-cap marginally increases the investment in cost reduction. The increased incentive is caused by the fact that the lower the price-cap, the more units the firm is selling (since price is lower); thus, 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 where the firm does not know the decline in marginal cost that will result from a given level of cost-reducing effort, the results are not as clear. The authors are able to show that in some

the firm typically has 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, for example, to rural areas, continue in a different form: telecommunications firms are taxed and the proceeds are used for direct subsidization, rather than accomplishing subsidization via the price structure.

20 The lower the price-cap, the lower the ceiling on the price the regulated firm can charge.
In summary, Cabral and Riordan show that a firm regulated under a price-cap mechanism will have an incentive to invest in cost reduction, as long as the price-cap is not too low.

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. Cabral and Riordan’s basic results are confirmed: in a model where it is known when price-cap regulation will end (and regulation will revert to rate of return), 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 a model where price-cap regulation continues indefinitely. Clemenz also shows that price-cap regulation is capable not only of raising welfare in general but also of increasing consumer surplus.

Sibley (1989) and Lewis and Sappington (1989) derive the optimal regulatory policy in a given situation where the regulated firm and the regulator have different information; the optimal regulatory mechanism is similar to price-cap regulation. Sibley assumes the regulated firm has full information about technology and demand, while regulators have no information; he also assumes that regulators can observe lagged prices, outputs, and expenditures by the firm. The regulatory mechanism is derived in such a way that it leads to efficient pricing and input usage; this mechanism is remarkably similar to a price-cap regime. Lewis and Sappington present a model in which the regulated firm has full information about its capabilities and cost-reducing activities. The 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 from cost reduction with consumers). They find 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.

### 3.3 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 impact on quality, the impact of a finite time period for price-cap regulation, and issues regarding re-negotiation of the terms of the regulation.

Typically, the price-cap is implemented as requiring that a price index of the different products provided by the regulated firm be below the price-cap;
at issue are the weights used to form the price index. Law (1993) shows that using a Laspeyres index in 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:

$$\sum_i \left( \frac{p^t_i}{p^t_{i-1}} \right) \times \left( \frac{\sum \left( \frac{p^t_i x^t_i}{\sum p^t_{i-1} x^t_{i-1}} \right) \leq 1 + CPI - x, \right)$$

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 is the weight, and 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 price-cap set tomorrow via the weight, 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. 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. Foreman shows that replacing the revenue weights with quantity weights reduces the effect of weight manipulation, especially as demand becomes less elastic.21

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) considers the effect of price-cap regulation on reliability of supply in electric supply; 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

---

21 Further discussion of index issues is contained in Cowan (1997), Diewert and Fox (2000), and Vickers and Yarrow (1988).
the problem of lower reliability when the price-cap is binding, but introduces a new problem of overpricing 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 re-negotiating 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. For example, 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 and “previously unsuspected financial strength”. Thus, the commitment not to intervene for five years did not last even 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.

Armstrong, Rees, and Vickers (1995) derive a model of price-cap regulation with an endogenously determined length of time until the next review. They identify the following trade-off. The longer the amount of time between reviews, the more likely it is that the price-cap will be set at a low level at the next review. This occurs because, prior to the review, firms retain profits for a longer period and thus have a higher pay-off to investing in cost reduction. However, the cost of the higher incentive for cost reduction comes at the expense of a longer period of time during which price is high. They also demonstrate that demand elasticity and the effectiveness of investments in cost reduction are key determinants of the optimal interval between hearings. Coco and De Vincenti (2008) examine whether a price-cap review improves welfare and, if it does, the optimal number and timing of reviews. They find that a review can be justified on purely productive efficiency grounds. The trade-off is between (a) reducing cost-saving effort in the early period, because any increased profits to be captured from cost savings are kept by the firm for a shorter time period (up until the review) and (b) increasing cost-saving effort

Coco and De Vincenti’s model is equally one of cost-based regulation with regulatory lag. They assume that the regulator observes past cost and knows demand, and at each price review sets the price-cap equal to average cost of production. Thus, their analysis is applicable to the optimal length of review under rate-of-return regulation where the allowed return is equal to the firm’s cost of capital.

---

22 Coco and De Vincenti’s model is equally one of cost-based regulation with regulatory lag. They assume that the regulator observes past cost and knows demand, and at each price review sets the price-cap equal to average cost of production. Thus, their analysis is applicable to the optimal length of review under rate-of-return regulation where the allowed return is equal to the firm’s cost of capital.
in the future period (after the review), because the review will lead to an increase in output to which the future cost-saving effort will apply. Their model is one of transparency; given sufficiently price-responsive demand and sufficiently low disutility of cost-saving effort, welfare is improved under a price review that the regulated firm correctly foresees.

Weisman (1993) theoretically analyzes the effect of re-negotiation of the terms of the price-cap regulation regime on firm behavior. Given a non-zero probability that the regulator will re-write 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.

Isaac (1991) considers the end-game problems that may arise when a firm is regulated under a price-cap for some known, finite period of time, after which rate-of-return regulation will be implemented. 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 to use the rate review when moving back to rate-of-return regulation to appropriate profits earned by the firm during the price-cap regime.

3.4 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. However, 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.

In practice, price caps are often put in place in conjunction with earnings-sharing schemes. For example, firms may be required to return a portion of their earnings above a certain amount to consumers. 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.”

---

23 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.
provisions, the firm is allowed pricing flexibility below the price-cap, provided that the rate of return is not above its cap. 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 is that such an approach eliminates one benefit of 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 (1996) 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. In addition, 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.

### 3.5 Theoretical Comparison of Incentive-Based Regulation to Other Forms of Price Regulation

Cabral and Riordan (1989) derive models of price-cap regulation and then compare the outcomes to several alternative regulatory regimes. Cabral and Riordan 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; the firm keeps any increase in profits between the time of cost reduction and the time of the new rate hearing. However, the longer regulatory lag and the price flexibility in a price-cap regulatory regime further increase the incentive to reduce costs. Cabral and Riordan also compare the level of price under the two alternative regulatory regimes. 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 and rate-of-return regulation. She points out that, under price-cap regula-

---

24 In some states, adjustments may be made if the firm’s rate of return falls below a floor, as well.
tion as implemented in the UK, hearings occur at fixed intervals and use all information available since the previous hearing. On the contrary, 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 are 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 overinvestment in capital and underinvestment in effort, especially during periods when cost data are gathered for hearings. Pint 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 if 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, welfare under rate-of-return regulation can be increased by incorporating two aspects of price-cap regulation: fixed intervals between hearings and using all data since the previous hearing.

Several papers use simulation techniques to compare regulatory regimes and the degree to which welfare may be increased. Schmalensee (1989) 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 $P = \rho + \gamma(C - \alpha)$, where $P$ is the regulated price, $\rho$ is the base price, $\gamma$ the cost-sharing fraction, $C$ is observed cost, and $\alpha$ is the expected unit cost under the pre-reform regulatory regime. In this formulation, 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 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 (1989) 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.

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 completely replace cost-based methods, and thus new problems arise. Whether incentive-based regulation is superior to cost-based regulation in practice is an empirical inquiry.

3.6 Empirical Analyses 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. 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.

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 efficiency, but only with a lag; price-cap regulation in conjunction with profit-sharing leads to a smaller but immediate impact on efficiency; and a pure profits-sharing scheme actually harms efficiency. Mathios and Rogers (1989) find that prices in states that regulated AT&T using a price-cap are 7–10% lower than in states that use rate-of-return regulation. They use variation in state regulatory mechanisms with a simple price-cap/no 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.

Several studies (see, e.g., Armstrong et al. 1994 and Price 1992) examine the move to price-cap regulation in the UK. Arocena and Price (2002) point out that in the UK the move to price-cap regulation coincided with privatization; as such, comparisons before and during price-cap regulation are confounded with the simultaneous move from public to private ownership. They therefore examine the switch from cost-based to incentive-based regulation in Spain, where ownership structure did not change; about half the electric generation firms were public and about half private. They find that public

---

25 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.
sector generators are on average more efficient than private sector generators under cost-based regulation, but that the private sector becomes more efficient than the public sector under price-cap regulation. Price-cap regulation had a stronger efficiency effect on private sector generators than public sector generators. Arocena and Price close by observing that they find that private sector generators in Spain improved efficiency by 1.9% a year under the price-cap, compared to an annual increase in the UK of 5.6% under the introduction of a price cap and privatization.

Hattori, Jamasb, and Pollitt (2005) compare productivity growth for electricity distribution under rate-of-return regulation (in Japan) and under price-cap regulation (in the UK) using both Data Envelopment Analysis (DEA) and Stochastic Frontier Analysis (SFA). “DEA is a non-parametric method and uses piecewise linear programming to calculate (rather than estimate) a sample’s efficient or best-practice frontier”, whereas SFA “is a parametric method used to estimate the efficient frontier”.26 They find that UK firms are generally more efficient than Japanese firms, although the gap is narrower when operating expenses rather than total expenses are used as a measure of inputs. The results also suggest that the firms that are most efficient in the UK have increased their lead over lagging firms during price-cap regulation.

Eckenrod (2006) examines the effect of the transition from rate-of-return regulation to price-cap regulation in telecommunications in the US on prices and markups. She finds that while markups did increase (to the benefit of firms), prices also fell; thus, the move to price-cap regulation in telecommunications seems to have benefited both firms and consumers.

Overall, empirical work suggests that incentive regulation as implemented has improved welfare relative to rate-of-return type regulation.

4. REGULATION OF NETWORKS

4.1 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 local network access from

---

26 Hattori, Jamasb, and Pollitt (2005), pp. 28 and 31, respectively.
the regulated local phone company on which the competitor can send calls. Under this structure, the central 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. 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.

Networks are relevant to telecommunications, electricity, and natural gas. For telecommunications, relevant networks are the local fixed network; local mobile network; and long-distance networks. Telephone service is provided over these networks, and connecting the networks is necessary to complete calls between networks (e.g., for a cell phone to call a landline or an AT&T landline to call a Qwest landline). For electricity, the relevant networks are for transmission (done at high voltage) and distribution (done at low voltage). Increasingly, competitive generation and retail (or marketing) sectors provide electricity generation (the input) and retail sales (the output) over these networks. Often the industries involved regulation of vertically integrated firms. As technology changed and economic learning advanced, efforts to separate the competitive stages of the industries (e.g., long-distance phone service and electricity generation) occurred. For example, by 2005, competition in electricity generation at the wholesale level was complete in many European countries (Jamasb and Pollitt 2005).

Why is regulation necessary in this situation? The economic justification continues to be to reduce the deadweight loss that arises under monopoly. Often a physical network leads to a situation of natural monopoly, where it is cheaper to have the physical network provided by a single firm. However, the issue is more complicated because the owner of the network typically competes with others in sectors that are potentially competitive. A natural

---

27 Typically, provision of the network remains a regulated monopoly due to economies of scale.
28 Specifically, the countries include Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden, and the UK.
29 This is not necessary, however. For example, the provision of a mobile phone network, even though it requires physical facilities (cell phone towers), is not a natural monopoly. Likewise, cost conditions in long-distance telephone service do not support a natural monopoly. See Armstrong (1997).
30 When AT&T was broken up, the initial industry structure prohibited the regional operating companies (local service) from competing in the potentially competitive long-distance market. With the Telecommunications Act of 1996 and deregulation of electricity in the US, the exclusion of network owners from the competitive sectors is declining.
incentive arises for network owners to price access to the network in such a way as to give itself an advantage in the competitive sectors. This is relevant for the case both of one-sided access, where one firm provides access to its network, and two-sided access (sometimes called interconnection), where two networks are connected. 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.

4.2 One-Sided Access or, Simply, Access

One common industry structure is when the incumbent firm controls the physical network that is a necessary input into the downstream product, where the incumbent faces competition. Then providers of the competitive, downstream product require access to the network as an input to providing the service. At issue are not only access to the network itself, but the terms of the access (pricing and quality). Access requirements and charges are signals that affect downstream entry, upstream bypass, network investment, and network utilization (Vogelsang 2003).

The socially optimal access price and retail prices could be solved à la Ramsey – social welfare could be maximized subject to the incumbent’s budget constraint, assuming that provision of the network is characterized by economies of scale.31 Vogelsang (2003) emphasizes the tremendous informational requirements to achieve the optimal prices, as “they incorporate the incumbent’s budget constraint, demand relationships, cost relationships, and types of competition”. He reports that no regulator has tried to implement such pricing. Access pricing in practice largely falls into three categories: (1) access prices set based on costs; (2) access prices set according to the efficient component-pricing rule (ECPR); or (3) access prices subject to a price cap (which may apply to access only, or may apply to access and the price for the retail service).

Cost-based access charges An approach to access that is used generally in the UK and most of the EU is to price access based on incremental costs.

31 Ramsey (1927) introduced a 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. Thus, Ramsey or Ramsey-Boiteux prices form a price structure that is optimal in that it has the lowest deadweight loss given the need to cover the total costs of the firm.
Incremental costs may be significantly more than marginal costs for large changes in output.

One version of cost-based access charges is described by Baumol and Sidak (1994). Suppose that a regulated natural monopoly earns $T$ from a particular service towards its overall revenue shortfall, defined as the difference between total incremental revenue less total incremental cost. Assuming 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 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 that of the regulated firm. Thus, in practice, access charges are generally increased above incremental costs to cover common fixed costs.

Efficient component-pricing rule (ECPR) access charges 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.32 This approach was ultimately overturned in New Zealand, but it is commonly used in the US to determine the access of competitive local exchange carriers (CLECs) to the incumbent local exchange carrier’s (ILEC’s) network.33

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

---

32 The rule is generally attributed to Willig (1979) and Baumol (1983) and is extensively discussed by Baumol and Sidak (1994). 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 during the 1990s.

33 Vogelsang (2003).
out an additional potential benefit. By including the regulated firm’s opportunity cost in the access charge, subsidies mandated by regulators can be maintained. 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. Entry into the high-profit markets can make the regulatory regime unstable, but 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 (1994) 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. They also show that under different forms of imperfect 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.” For example, whether the products are homogeneous or differentiated and whether they are produced subject to constant or increasing returns to scale are important determinants of final prices, whether the final prices are set by the market or by regulators. It seems clear that the ECPR does not have the wide applicability first claimed.

---

34 I call this a potential benefit since many of these cross-subsidies are not efficient to begin with.
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 ECPR.

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 relaxed 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. However, they also show that, when incorporating the firm’s budget constraint, a positive Ramsey 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. Finally, they 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, since 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 by 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: various 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 entrant, various regulatory regimes, and under the possibility of bypass. They also compare the optimal access price with 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 fails 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 pricing with interdependent demands, but also show that ECPR is not always efficient. In particular, in order for a Ramsey 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 zero elasticity of substitution 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) competition in prices ensues in the downstream market, which in combination with assumption (5) result 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.

Access charges subject to price-caps Laffont and Tirole (1996) consider a completely different approach to regulating interconnection. They suggest that access could be regulated under a global price-cap. Under this approach, the price-cap would apply to 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 of this solution is that it requires no more informational requirements than existing regulatory policies. The regulated firm will adopt the optimal Ramsey price structure given its information about demand and cost; the regulator does not require information on marginal costs nor demand elasticities. 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.

\[ \text{(footnote 31)} \]

\[ \text{Ramsey prices obtain so long as all goods, including access, are included in the definition of the price cap and the weights are exogenously set at the level of output to be realized.} \]
as shown by Armstrong, Doyle, and Vickers (1998). The benefits identified by Laffont and Tirole of 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.

However, at least as of 2003, no regulator had ever used global price caps (Vogelsang 2003). In practice, separate price caps for access and retail services (introduced in the US in 1991 and in the UK in 1997, according to Vogelsang) have been implemented in telecommunications. Separate price-caps were implemented to reduce the incentive for incumbent firms to disadvantage their downstream rivals (discussed in Section 5 below). However, separating the price caps does not lead to socially optimal access and retail prices.

4.3 Two-sided Access Pricing or Interconnection

In some situations, the interconnection of two competing networks provides additional value to consumers via network externalities. This is common in telecommunications, including the case for competing long-distance, mobile, and local networks, and between these types of networks as well (e.g., a mobile phone calling a land-line). In these cases, there is significant demand for subscribers on one network to call subscribers on the other network. In other words, network externalities are important; the value to me of accessing the phone system increases the more people who are also connected to the phone system (the more people I can call and who can call me).

Armstrong (1997) provides a simple model of network interconnection that demonstrates a justification for regulation of access prices even when costs are such that competing networks are feasible. In the simplest setting, where two symmetric, competing firms can choose between a high and a low price, the firms have an incentive to choose the high access price. Essentially, the interconnection of networks provides a mechanism for the firms to “collude”. Although each firm is negotiating access charges independently, their interests coincide because the access charge is both an input cost (for completion of calls to the other network) and a source of profit (for completion of calls from the other network). The maximization of social welfare calls for a low access price, which a regulator can implement. The tendency towards this form of “collusion” is reduced as the networks become asymmetric. The firm with a network that originates more calls than it terminates will want a low access charge while the other firm will want a high access charge.
5. REGULATION AND COMPETITION (ANTITRUST) LAW

5.1 Introduction to the Issues

There have long been two interventions by the government in the marketplace to sustain competition. For those industries where competition generally could be sustained by the market (i.e., non-natural monopoly cost structures), most governments have established competition (antitrust) laws to ensure that companies compete “fairly.” For those industries that are natural monopolies, governments have often regulated the firms in the industry to control the exercise of market power. As more portions of regulated industries have been deregulated in the last couple of decades, potential conflicts between regulation and competition (antitrust) law have arisen. These issues generally relate to the access issues described in Section 4.2 above.

The overlap between regulation and competition law typically involves a firm that controls a network that is an important input into a service provided by the network firm and its rivals. Consider the issue of access, where one firm controls a network that is necessary to provide a downstream service in an unregulated situation. The network monopolist makes the same profit if it monopolizes the downstream market (excludes all downstream rivals) or if the downstream industry is perfectly competitive. If the downstream industry is not perfectly competitive but the product is homogeneous, then the network monopolist and society are better off if the network monopolist excludes competitors, because doing so avoids the problem of double marginalization, where the input (network) monopolist increases price for the network input above cost, and then the imperfectly competitive downstream firms increase price for the output above the inflated price they are paying for the network input.

The presence of regulation can alter these incentives, however. Suppose that the firm that provides network services, which is a natural monopoly, is regulated in that area. The firm also provides products delivered via the network (e.g., voice telephony over the local loop) in competition with other

---

37 Another commonly used alternative was to provide the natural monopoly product or service via a publicly owned institution. This approach is falling out of favor because public firms are often thought to operate inefficiently. (See, however, Arocena and Price (2002) for an alternative view that public firms will operate more efficiently than regulated private firms.)

38 If the downstream product is differentiated, the social welfare evaluation is complicated; while exclusion of downstream firms eliminates double marginalization, it may negatively (or positively) affect variety.
firms. If the downstream market is not perfectly competitive, so that there are profits to be made, the incentive is clear – the regulated network firm would increase its profits if it could advantage itself in the downstream market. Regulation prevents the firm from extracting monopoly rents from its control of the network input, so the firm chases profits in the market in which it is not regulated. Thus, the presence of regulation can cause competition (antitrust) issues. Indeed, Mandy (2000) shows that, in the extreme, where the regulated price of the network input is at marginal cost, the incumbent has an incentive to completely foreclose downstream rivals, because the incumbent makes profits only from sales of the downstream service, not from sales of access to the network.

5.2 Using Network Market Power to Harm Competition

Consider an incumbent, regulated firm that controls access to a network that is a necessary input into the provision of service downstream. Examples include firms that control the local telephone service network and the electricity distribution system. Typically, these firms also provide the downstream service, and recently in the US and the UK they face competition in providing the downstream services. Then downstream rivals must either purchase the network input from the incumbent or build its own network. As long as the network is subject to sufficient economies of scale as to be a natural monopoly, welfare is harmed by the duplication of the physical network, and it is better to require the incumbent to provide access to the network.

When the price of the network input is regulated, the incumbent may resort to non-price means to disadvantage its downstream rivals. In the literature, these non-price means are typically referred to as sabotage. Mandy (2000) lays out the intuition. The incumbent makes profits from selling the network input and from selling the downstream service. Sabotage increases the marginal cost of downstream rivals. This reduces the amount of input the rivals purchase, reducing the incumbent’s profit from sales of the input. However, it increases the incumbent’s profit from sales of the downstream service, of which the incumbent now supplies more. On the other hand, if the rivals are more efficient in providing the downstream service, of which the incumbent now supplies more. On the other hand, if the rivals are more efficient in providing the downstream service, the incumbent may be better off not sabotaging the downstream rivals. Thus, depending on parameter values, the incumbent may increase profits by sabotaging downstream rivals.

5.3 The Relationship between Networks and Competitors

One way to deal with the incentive for the network operator to disadvantage rivals in competitive upstream or downstream segments is to separate the
activities. This is typically done in one of four ways: functional (or management), accounting, legal, or ownership separation. The latter is the most effective, as the remaining network firm has no vested interest in upstream or downstream activities. Jamasb and Pollitt (2005) report that among European countries, about half require legal separation of electricity transmission system operators from generation and retailing and about half require ownership separation; about half require legal separation of distribution system operators from generation and retailing, a quarter require accounting separation, and about a quarter require management separation.

Vertical separation, when effective, can eliminate the incentives of the regulated network incumbent to favor upstream and downstream firms. However, if there are economies of scope in providing the network input and the upstream or downstream product, the cost of vertical separation will be a loss of such economies. In this case, regulation of the network incumbent (or the separated upstream or downstream firms) could not recover these economies.

5.4 The Overlap of Regulation and Competition (Antitrust) Law

Perhaps the seminal case in the US covering the interplay between regulation and competition law is Verizon Communications v. Law Offices of Curtis V. Trinko (hereafter, Trinko). This case involved Verizon’s obligation, under the US Telecommunications Act of 1996, to provide unbundled network elements (UNEs) to competitive local exchange carriers, including AT&T. The relationship between Verizon and AT&T is subject to regulation at both the state and Federal levels. The New York Public Service Commission (PSC) approved the agreement between Verizon and AT&T regarding Verizon’s supply of UNEs to AT&T. Under the terms of the Telecommunications Act of 1996, the FCC allowed Verizon to enter the long-distance market after finding that Verizon provided UNEs to its rivals for local phone service, including AT&T. AT&T and other competitive local exchange carriers complained that Verizon was not adequately supplying UNEs, and the PSC and FCC opened investigations and imposed penalties and behavioral changes. Subsequently, the Law Offices of Curtis V. Trinko, a customer of AT&T, filed suit alleging that Verizon’s actions were an attempt to monopolize local telephone service in violation of Section 2 of the Sherman Antitrust Act.

The provision of the Telecommunications Act of 1996 that allowed local network incumbents to enter the long-distance market only after sufficient competition existed in the local market was designed to counteract the local network incumbents’ incentive to foreclose entry into the market for local calls.
The claim against Verizon is of the type described in Section 5.2: Verizon allegedly filled its competitors’ orders in such a way as to discourage customers from signing up or remaining with its rivals’, including AT&T. Thus, Verizon’s actions allegedly blocked entry and harmed rivals’ ability to compete in the market for local telephone service. The Supreme Court considered “whether a complaint alleging breach of the incumbent’s duty under the 1996 Act to share its network with competitors states a claim under §2 of the Sherman Act” (2004 WL 51011 (US)).

The Supreme Court’s ruling laid out the extent to which competition law or regulatory policy takes precedence. In particular, the Telecommunications Act of 1996 specifically stated “nothing in this Act … shall be construed to modify, impair, or supersede the applicability of any of the antitrust laws” (47 USC §152). The Supreme Court also found that the Telecommunications Act of 1996 did not create new antitrust claims; “[t]hat Congress created these duties [to provide UNEs to rivals] does not automatically lead to the conclusion that they can be enforced by means of an antitrust claim” (2004 WL 51011 (US)). The Supreme Court ruled that the conduct of Verizon did not violate antitrust case law.

However, the Supreme Court did state that the regulatory regime was an important consideration in analyzing the competitive impact of the defendant’s conduct, and in particular noted that the regulatory regime itself was effective in maintaining competition: “The regulatory framework that exists in this case demonstrates how, in certain circumstances, regulation significantly diminishes the likelihood of major antitrust harm” (2004 WL 51011 (US)).

Blair and Piette (2005) identify the crucial issue in the Trinko ruling: if the US antitrust laws cannot be used to enforce the Telecommunications Act of 1996, will its provisions be effective? Via the Telecommunications Act of 1996, Congress seemingly decided that the local telephone service market should be competitive by requiring incumbent local exchange carriers (ILECs) to provide access to their facilities to competitive local exchange carriers (CLECs). The Telecommunications Act of 1996 sets the terms of access, and regulatory agencies are available to implement these terms. For example, the problems with access alleged by AT&T (the CLEC) against Verizon (the ILEC) were investigated by both the New York Public Service Commission and the FCC. Both imposed fines against Verizon (the state agency ordered the fines to be paid to the CLECs, while the FCC’s action involved Verizon’s “contribution” to the US Treasury) for a total of $13 million. Blair and Piette note that these regulatory actions may not be sufficient to deter the ILECs from foreclosing entry into the local market. To be a sufficient deterrent, the expected fines must be sufficient to make interference in access unprofitable. While the FCC and state regulatory authorities can impose fines, they are legally limited in the magnitude of the fines. In Blair and Piette’s view, there...
is ample evidence that ILECs have interfered with CLECs’ access to the network, and thus it appears that the regulatory bodies are not sufficient to effectively implement the Telecommunications Act of 1996.

In the EU, the Court of First Instance has ruled that a company is not cleared of complying with competition law even if its prices must be approved by a regulatory body. Deutsche Telekom had been accused of abusing its monopoly power by overcharging its rivals for access to local telephone loops; in particular, it was alleged that Deutsche Telekom charged higher access fees to its rival telephone service providers than it did to its retail customers. Again, this action is similar to the conduct described in Section 5.2: Deutsche Telekom controls access to the physical network, which it and its rivals need to provide telephone service to retail customers. By charging higher prices to its rivals than to its customers, Deutsche Telekom would be able to disadvantage its rivals and monopolize the market for local telephone service. Deutsche Telekom argued that the European Commission was encroaching on the German telecommunications regulator’s (RegTP’s) authority; the Court disagreed.

6. CONCLUSION: REGULATION IN PRACTICE

As regulation theory has advanced, especially in the last 30 years, 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. In particular, deregulation occurred in railroad and trucking, air passenger transportation, natural gas, electricity, telecommunications, and cable television. Other countries have seen an impressive move away from handling natural monopolies by nationalizing them to regulation of privatized firms. These countries, especially the UK, 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. There is a general movement away from cost-based regulation towards incentive-based regulation in an effort to achieve a better outcome. The UK led the way with adoption of price-cap regulation in 1984,

---

40 For example, regulation of trucking in the US largely developed in order to protect regulation of railroads, which were thought to be natural monopolies.

41 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. However, cable was re-deregulated with the passage of the Telecommunications Act of 1996.
after privatizing its telecommunications industry. Privatization of other industries spread the use of price-cap regulation to electricity, gas, and water. In practice, price-cap regulation in the UK 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.

Technological advances, as well as the move towards incentive-based regulation, have highlighted two, related, issues: the regulation of network industries and the interplay between regulation and competition law. Networks are often necessary inputs into service provision, either on a one-sided basis (e.g., electricity generators must access transmission and distribution networks to provide electricity to users) or on a two-sided basis (e.g., two telephony networks (be it for local, long distance, or mobile service) interconnect to allow users of one network to reach users of the other network). Regulation of access or interconnection rates must be designed with the market structure of the network and the connected downstream or upstream firms or the connected networks. The incentives for a network firm to interfere with competition in downstream or upstream markets calls competition law into play. What is allowed under regulation may negatively impact competition in other markets, and these interactions can sometimes lead to ineffective regulation and competition law (which perhaps describes the US) or contradictory demands of regulation and competition law (which perhaps describes the EU).

BIBLIOGRAPHY

Price regulation


