REGULATION, COMPETITION AND LIBERALIZATION

by

Mark Armstrong* and David E. M. Sappington**

ABSTRACT

In many countries throughout the world, regulators are struggling to determine whether and how to introduce competition into regulated industries. This essay examines the complexities involved in the liberalization process. While stressing the importance of case-specific analyses, this essay distinguishes liberalization policies that generally are pro-competitive from corresponding anti-competitive liberalization policies.

September 2005

* University College London.
** University of Florida.

We are grateful to Sanford Berg, Severin Borenstein, Cory Davidson, Roger Gordon, Antonio Estache, Jerry Hausman, Mark Jamison, Paul Joskow, Mircea Marcu, John McMillan, Paul Sotkiewicz, John Vickers, Catherine Waddams, Helen Weeds, and anonymous referees for helpful comments and observations.
1. Introduction.

Economists have developed an extensive set of principles for regulating a monopoly supplier. The benefits of unfettered, pervasive competition are also well documented and well understood. However, our understanding of the precise conditions under which regulated monopoly supply is preferable to unregulated competition is limited. Furthermore, we know relatively little about optimal liberalization policies – the policies that govern the transition to competitive market conditions – in cases where competition is deemed superior to monopoly.

The purpose of this essay is to explore these two issues, both of which are of substantial practical importance throughout the world. These issues are particularly relevant in key network industries (such as the telecommunications, natural gas, electricity, transport, and water industries) where scale economies can render production by many firms uneconomic, but where some competition may be useful to help discipline incumbent suppliers of key services. Our analysis of the choice between regulated monopoly and unregulated competition, like our analysis of the design of liberalization policies, emphasizes the problems that imperfect information and imperfect institutions pose for the design of industry policy.

This essay has three main parts: (1) section 2 reviews some recent experience with liberalization policy; (2) sections 3 through 5 consider the choice between unregulated competition and regulated monopoly; and (3) sections 6 and 7 consider the design of liberalization policies in settings where competition is preferred to monopoly.

Section 2 summarizes selected experiences with liberalization in three network industries where liberalization has garnered significant attention in recent years: telecommunications, natural gas, and electricity. The experiences in Chile’s telecommunications industry, the UK’s natural gas industry, and the state of California’s electricity industry have varied greatly, ranging from substantial success in Chile to substantial failure in California.

Section 3 begins the discussion of the choice between unregulated competition and regulated monopoly by analyzing simple formal models of these two forms of industrial organization. The models emphasize the role of imperfect information. Section 4 considers additional factors that affect the choice between regulation and competition, including the resources available to the regulator, the regulator’s independence and autonomy, the need to ensure ubiquitous service at affordable prices, and the importance of investment and innovation. Section 5 notes that even when monopoly supply is the preferred mode of industry operation, some of the benefits of competition may be secured by allowing potential operators to bid for the right to serve as a regulated monopoly supplier.
Section 6 begins the discussion of liberalization policies in settings where the benefits of competition outweigh its costs. The potential merits and risks of direct entry assistance are considered first. Liberalization policies that are likely to reduce the intensity of long-term industry competition (and therefore generally are not recommended) are reviewed next. These policies include providing temporary monopolies or oligopolies, excluding foreign investors, specifying market share targets for industry participants, restricting incumbent suppliers asymmetrically, and affording competitors unduly favorable long-term access to the incumbent’s infrastructure. Section 7 discusses liberalization policies that typically will enhance the intensity of long-term industry competition (and therefore are recommended). These policies include reducing the costs that customers incur when they switch suppliers, re-balancing the prices charged by the incumbent supplier to better reflect its operating costs and otherwise re-designing the regulations imposed on the incumbent supplier to account explicitly for emerging competition, privatizing state-owned enterprises, establishing appropriate access prices, and increasing monitoring, data reporting, and antitrust scrutiny, at least during the early stages of liberalization.

Two central themes emerge in this essay. First, even the comparatively simple choice between regulated monopoly and unregulated competition can be intricate and complex in practice. Second, the decision to introduce competition into an industry is only the beginning of a journey down a long and winding road that can present many obstacles and detours. Furthermore, the best route from monopoly to competition can differ substantially in different settings. Therefore, there is no single set of directions that can guide the challenging journey from monopoly to competition in all settings.

Even though detailed, comprehensive directions typically are not available, some broad conclusions (summarized in section 8) that can serve as useful guide posts can be drawn. These broad conclusions include the following five. First, the greatest potential gains from competition tend to arise when: (i) industry scale economies are limited relative to consumer demand; (ii) the industry regulator has limited information, limited resources, and limited instruments with which to craft policy; (iii) the regulator’s commitment powers are limited; and (iv) subsidization of the consumption of some of the dominant supplier’s services either is not critical or can be achieved by means other than through distortions in the supplier’s price structure.

Second, there are a wide variety of liberalization policies, and the merits of the different policies vary considerably. Therefore, it generally is more appropriate to inquire about the benefits and costs of specific liberalization policies than to ask whether liberalization per se is desirable or undesirable. Third, liberalization policies that primarily aid some competitors and handicap others
on an ongoing basis can hinder the development of vigorous long-term competition. Fourth, liberalization policies that remove entry barriers and empower consumers to discipline industry suppliers typically are the best means by which to foster vigorous long-term competition. Fifth, well-designed liberalization policies can facilitate a transition from stringent, detailed regulatory control to less intrusive antitrust oversight. However, during the transition process, heightened regulatory and antitrust scrutiny may be required.

2. Recent Experience with Liberalization.

To help ground the ensuing discussion of the principles of liberalization policy, we begin by reviewing briefly the recent experience with liberalization in Chile’s telecommunications industry, the UK’s natural gas industry, and California’s electricity industry.

A. Liberalization in Chile’s Telecommunications Industry.\(^1\)

The General Law of Telecommunications opened Chile’s telecommunications industry to competition in 1982. This law generally did not restrict the number of licenses that would be granted to deliver telecommunications services in Chile. Importantly, the licenses were for non-exclusive provision of wireline and wireless services, both local and long distance. The primary obligation imposed on all telecommunications operators in Chile was to connect their networks in compliance with specified technical requirements. Other terms and conditions of interconnection typically were left to negotiation among the operators.

In 1982, local telecommunications services were supplied almost exclusively by Compañía de Teléfonos Chile (CTC) while long distance telecommunications services were supplied almost exclusively by Entel. Not surprisingly, these two state owned enterprises were not anxious to connect their networks with the networks of emerging rivals. Few interconnection agreements were signed and competition was limited until Chile implemented a dispute resolution process that ensured the timely execution of interconnection agreements. Long distance competition also was limited until an equal access requirement was imposed in 1994. The requirement stipulated that every time a customer placed a long distance call, she had to explicitly designate (via dialing a special code) a long distance company to carry the call. The special code for each long distance carrier – incumbent and entrant alike – had the same number of digits. Most importantly, a call was not automatically carried by Entel if the caller did not specify an alternative carrier.

---

Once new suppliers were afforded substantial opportunity to provide telecommunications services on terms comparable to those faced by incumbent suppliers, competition began to flourish in Chile. The number of fixed lines more than tripled (from roughly one million to more than three million) between 1992 and 2000, and the extensive waiting list for (fixed) wireline telephone service that had inconvenienced Chile’s citizens for so many years quickly disappeared. The number of mobile telephone subscribers in Chile increased nearly ten-fold (from roughly 36,000 to more than 3.4 million) during the 1991-2000 period, and nearly doubled again (to more than 6.7 million) between 2000 and 2003. Prices for most telecommunications services in Chile also have declined substantially since 1990.

By 2003, nearly one-fourth of the fixed telecommunications lines in Chile were supplied by CTC’s competitors, rather than by the incumbent supplier. Competitors secured especially large market shares in Chile’s urban regions. This pattern of entry likely reflects in part the lower unit costs of serving urban areas, given their relatively high population densities. This entry pattern also may reflect regulated prices of basic telephone service that are further above costs in urban regions than in rural regions. The regulated prices of local telephone services in rural regions often do not reflect fully the relatively high unit costs of serving these regions. Instead, prices in rural regions tend to be set closer to (and sometimes below) cost while prices in urban regions often are set substantially in excess of cost. Such pricing patterns are designed to ensure that basic telephone service is affordable to all citizens. These patterns may be less pronounced in Chile than in some countries because a separate fund has been established to subsidize the purchase of basic telephone service in rural regions and low-income urban regions. Nevertheless, in jurisdictions where CTC is deemed to be a dominant supplier of local telephone services, it is required to set the same price for basic services across substantial geographic regions. Some suggest that this restriction on CTC’s ability to target price reductions to those regions in which competition is most intense may be one cause of the substantial market share that competitive suppliers of local telecommunications services have been able to secure in Chile.²

B. Liberalization in the UK’s Natural Gas Industry.³

Natural gas was one of the principal network industries in the UK targeted for privatization under Margaret Thatcher’s program of industrial reform. Natural gas was first extracted from the UK Continental Shelf in 1965, and British Gas (BG) was formed as a state-owned company in

---

² See Paredes (2005), for example.
³ The material is drawn from Armstrong et al. (1994, chapter 8) and Newbery (1999, chapter 8).
1972. BG held a legal monopoly over the sale of gas to consumers and a legal monopsony over the purchase of gas from producers in the UK fields. Since gas exploration and production involve substantial sunk costs, BG’s monopsony power necessitated the use of long-term (e.g., 25 year) purchase contracts to limit expropriation of the substantial investments made by gas producers.

An initial attempt to liberalize the gas sector occurred in 1982, when entrants were authorized to employ BG’s pipelines (at terms negotiated with BG) to supply gas to final customers. In addition, BG’s legal monopsony was removed. Although these steps were intended to facilitate competition in the gas industry, competition did not emerge immediately.

BG was privatized as a vertically integrated monopoly in 1986. At this time, consumers were divided into two categories for regulatory purposes: (1) the tariff market, consisting of those consumers who purchased less than 25,000 therms of gas per year; and (2) the contract market, consisting of the remaining higher-volume consumers. BG continued to enjoy a legal monopoly in the tariff market after privatization. BG’s prices in this sector were regulated: the annual increase in the average price per therm was limited to \( RPI + Y - X \), where \( RPI \) is the inflation rate, \( Y \) is a measure of the increase in the price BG pays for gas, and \( X \) is a productivity improvement factor.\(^5\) \( X \) was set at 2% in 1986, but raised to 5% in 1992.\(^6\)

Competition in the contract market was potentially open to competition, and BG’s prices in this market were not regulated. Although it was legal, competition did not emerge in this market until 1990. The delay reflected in part limited regulation of the prices BG charged retail competitors for access to its pipelines. BG also practiced price discrimination in the contract market, charging higher prices for gas to those industrial users who had no reasonable alternative energy source (e.g., electricity).

To combat these perceived problems, new regulations were imposed in 1988. The regulations required BG to: (1) publish a tariff on which any customer in the contract market could purchase gas (in an attempt to limit price discrimination); (2) publish a corresponding tariff of network access charges (although these charges remained otherwise unregulated); and (3) purchase no more than 90% of the gas in any newly discovered gas field. This final requirement essentially

\(^4\) Alternatively, BG could have been separated into a national pipeline business and several regional suppliers before privatization, an approach subsequently pursued in the UK electricity industry.

\(^5\) This \( Y \) factor enabled BG to pass through its costs of purchasing gas to consumers, and thereby limited BG’s incentive to secure low gas prices. Subsequent regulation permitted BG to pass through only a more exogenous index of gas costs.

\(^6\) The 2% value for \( X \) was widely believed to underestimate the realistic potential for productivity gains. Of course, lenient price regulations enhance revenues from privatization.
forced some limited entry into the contract market. By 1990, though, BG supplied 93% of the gas sold in this market.

Limited competition in the contract market promoted three further regulatory reforms in the early 1990s. First, the monopoly threshold in the tariff sector was reduced from 25,000 therms to 2,500 therms, thereby permitting smaller customers to purchase gas from competitors. Second, the prices charged for access to BG’s pipelines came under regulatory control. Third, BG was required to reduce to 40% (by volume) its share of sales to the contract market by 1995. At the same time, new gas fields were being opened, and entrants were able to buy supplies from these new fields.

Although regulation did not force BG to separate its pipeline and gas supply operations, intense regulatory scrutiny and onerous separate accounting requirements led BG to undertake such vertical divestiture voluntarily. Furthermore, the monopoly franchise threshold of 2,500 therms was removed entirely in 1998, so all consumers were permitted to purchase gas from suppliers other than BG. Competition intensified, and all retail price controls were removed in 2002. Only access charges continue to be regulated. Many consumers continue to purchase natural gas from BG, even though BG charges more than some competitors for what is largely a homogeneous product. This fact may suggest that customer switching costs or other causes of customer inertia are important in this industry.

C. Liberalization in California’s Electricity Industry

Although the state of California’s experience with liberalization in its electricity industry is quite recent, the experience is already legendary. In 1996, California enacted legislation that introduced five primary changes in the state’s electricity sector. First, electricity generation and wholesale prices for electricity were deregulated. Second, as they were required to do, the three incumbent (vertically integrated) suppliers of electricity sold a sizable portion of their generation capacity, focusing instead on the transmission and delivery of electricity. By 1999, the incumbent suppliers had sold to five independent suppliers generation capacity that produced roughly one-third of the state’s electricity consumption. Third, retail competition was introduced, as retail customers

---

7 These access charges were distance related. Because BG’s retail prices did not vary across geographic regions, this pricing policy enabled entrants to secure a higher profit margin when serving customers located close to the gas landing point.

8 As a result, BG was left with long-term obligations to purchase gas that it no longer needed to serve its own consumers. When BG was privatized, investors were promised that BG would be able to supply approximately two-thirds of the demand for natural gas by volume until 2009. As regulation and liberalization developed, this promise was not honored.

9 The following discussion is based primarily on Borenstein (2002).
were permitted to purchase electricity from firms other than the incumbent suppliers. Fourth, a ceiling was imposed on the retail price that incumbent suppliers charged for electricity. Fifth, the three incumbent suppliers were afforded strong financial incentives to buy and sell electricity through the California Power Exchange (Cal PX) during its first four years of operation. Cal PX was established in 1998 to run the day-ahead market for electricity in the state of California. Although the incumbent suppliers’ participation in Cal PX helped to ensure substantial short-term supply of and demand for electricity, the incumbents’ corresponding limited use of long-term contracts for the purchase or sale of electricity ultimately proved to be detrimental.

In the summer of 2000, unusually high temperatures and very dry conditions produced both a substantial increase in the demand for electricity and a significant reduction in the available supply of electricity from hydro-electric generating units in the state. When less efficient generating units were dispatched to meet the high demand for electricity, the wholesale price of electricity soared well above the ceiling that had been imposed on retail electricity prices. Consequently, incumbent suppliers incurred severe financial losses. The higher wholesale prices were not sufficient to augment the supply of electricity greatly, due to the long lead time necessary to construct new generating capacity and due to the particularly high costs faced by established producers of peak load capacity. These high costs reflected less efficient operating technologies, particularly high wholesale prices for natural gas, and expensive pollution permits.10

In 2001, the state of California adopted drastic measures to quell the crisis that had developed in its electricity industry. The state established itself as the wholesale buyer of power for the incumbent suppliers in response to the reluctance of other parties (including wholesale suppliers of electricity) to deal with the financially distressed incumbents. The state also raised substantially the prices incumbent suppliers could charge for electricity, especially to large retail customers. In addition, to prevent these customers from securing electricity from alternative suppliers at lower prices, the state terminated retail competition. In essence, California’s experiment with deregulation and liberalization had effectively ended.

3. Regulated Monopoly and Unregulated Competition.

In an economic paradise, where a regulator is omniscient, benevolent, and able to fulfill any promise he makes, competition cannot improve upon regulated monopoly. In such a paradise, the regulator will ensure the firm produces the ideal range of services at the lowest possible cost and

10 It has also been suggested that some of the few large suppliers of electricity may have exacerbated the problem by intentionally reducing their supply of electricity, thereby raising the market-clearing price of electricity, and hence the profits of electricity suppliers. (See Joskow and Kahn (2002), for example.)
will set welfare-maximizing prices for these services. Consequently, industry performance would not improve if an additional firm operated in this setting.

Of course, the real world differs markedly from this paradise. In practice, regulators invariably lack important information about the markets they oversee, and so will not be able to direct and control perfectly the activities of a monopoly producer. Because of its daily operation in the industry and its direct contact with consumers, the regulated firm will be better informed than the regulator about the demand for the regulated services it supplies, the minimum possible current cost of delivering the services, and the potential for less costly future provision. This information asymmetry generally gives rise to an unavoidable trade-off between rent and efficiency: the firm can be motivated to operate efficiently, but only if it is awarded substantial rent for doing so. In particular, the firm will operate at minimum cost and attempt to satisfy the needs and desires of customers only if it is awarded the full surplus that its activities generate.\(^{11}\) However, such a generous award to the regulated firm typically will provide it with significant rent, and thereby reduce the net benefits enjoyed by consumers. To limit the rent that accrues to the regulated firm, some inefficiency typically is tolerated.

To examine the optimal resolution of this trade-off and to examine the impact of limited information on the choice between regulated monopoly and unregulated competition, consider the following simple model.\(^{12}\)

### 3.1 A Simple Model of Regulated Monopoly.

Suppose that when regulated monopoly is implemented, the regulator faithfully pursues the social goal of maximizing the expected value of \(V + \alpha U\), where \(V\) denotes the surplus enjoyed by consumers, \(U\) is the firm’s rent, and \(\alpha \in [0, 1]\) is a parameter. Because \(\alpha \leq 1\), society values consumer welfare at least as highly as the welfare of shareholders, perhaps because consumers are less wealthy than shareholders or because many shareholders reside in other jurisdictions.

\(^{11}\) Loeb and Magat (1979) provide a formal statement and proof of this conclusion. Sharkey (1979) critiques Loeb and Magat’s analysis. Finsinger and Vogelsang (1981, 1982) and Sappington and Sibley (1988) suggest dynamic modifications of the Loeb and Magat policy that afford less rent to the regulated firm and/or can be implemented with less precise information about the surplus generated by the firm’s activities.

\(^{12}\) This model is drawn from the large literature that examines the design of regulatory policy when the regulated firm is better informed about its environment than the regulator. Reviews of this literature include Caillaud et al. (1988), Baron (1989), Laffont and Tirole (1993), Laffont (1994), and Armstrong and Sappington (2005). Joskow and Schmalensee (1986), Armstrong et al. (1994), Blackmon (1994), Sappington (1994), Mansell and Church (1995), Sappington and Weisman (1996), and Joskow (2005),
A transfer payment $T$ from consumers to the firm entails a reduction of $[1 + \lambda]T$ in the surplus enjoyed by consumers. The parameter $\lambda \geq 0$ represents the social cost of public funds. This cost arises from the distortions created by the taxes imposed on consumers/taxpayers to raise the funds.\(^{13}\) Notice that because public funds entail unit cost $1 + \lambda$, each dollar of public funds secured by taxing the profit of the regulated firm can be employed to increase consumer/taxpayer welfare by $1 + \lambda$ dollars.

The monopoly supplies a single product at regulated unit price $p \geq 0$. The demand curve for this product, $q(p)$, is common knowledge. The regulator sets both the unit price, $p$, for the regulated product and a transfer payment, $T$, from consumers to the regulated firm. The firm is obligated to serve all customer demand at the established price.

The firm incurs a fixed cost of operation, $F$, and a constant marginal cost of production, $c$. For simplicity, the firm’s marginal cost can take on one of two values, $c_L$ or $c_H$. Let $\Delta \equiv c_H - c_L > 0$ denote the difference between the high and the low marginal cost. The firm knows from the outset of its interaction with the regulator whether its marginal cost is high or low. The regulator does not share this information, and never observes the firm’s marginal cost directly. The regulator perceives the two possible cost realizations to be equally likely. For simplicity, the firm’s fixed cost of operation, $F$, is assumed to be common knowledge. The firm seeks to maximize its rent, $U$, which is the sum of its profit, $\pi \equiv q(p)[p - c] - F$, and the transfer payment, $T$, it receives from the regulator.

Suppose the regulator announces that he will set unit price $p_i$ and deliver transfer payment $T_i$ to the firm when the firm reports its marginal cost to be $c_i$, for $i \equiv L,H$. When the firm with cost $c_i$ chooses the $(p_i,T_i)$ option, its rent will be $U_i \equiv q(p_i)[p_i - c_i] - F + T_i$. The revelation principle ensures there is no loss of generality in examining regulatory policies that induce the firm to report its marginal cost truthfully.\(^{14}\) Therefore, social welfare when the firm’s marginal cost is $c_i$ is:

\(^{13}\) Laffont (2005, pp. 1-2) reports that for each dollar of tax revenue the government collects, citizens in a developed country bear a cost of approximately $1.30 (so $\lambda = 0.3$). The corresponding cost typically is substantially greater in developing countries.

\(^{14}\) See Myerson (1979), for example.
\( v(p_i) - [1 + \lambda]T_i + \alpha \left[ q(p_i)[p_i - c_i] - F + T_i \right] = w_i(p_i) - [1 + \lambda - \alpha]U_i, \)

where \( w_i(p_i) = v(p_i) + [1 + \lambda][q(p_i)[p_i - c_i] - F], \) and where \( v(p) \) denotes consumer surplus when price \( p \) is established. \( v(\cdot) \) is a convex function of \( p. \)

If the firm’s realized marginal cost were observed publicly, the regulator would implement the (Ramsey) price that maximizes \( w_i(\cdot) \) when cost \( c_i \) is realized, for \( i \equiv L, H \). These full-information prices will be marginal-cost prices \( (p_i = c_i) \) when \( \lambda \) is zero, because transfers from consumers to the firm entail no direct social costs in this case. In contrast, when \( \lambda \) is large, payments to consumers financed by the firm’s profit are highly valued, and the full-information prices will approximate the prices chosen by an unregulated profit-maximizing monopolist.

When the firm’s marginal cost is not observed publicly, it must be the case that

\[ q(p_L)[p_L - c_L] - F + T_L \geq q(p_H)[p_H - c_L] - F + T_H, \]

or, equivalently,

\[ U_L \geq U_H + \Delta q(p_H), \] (1)

to ensure the firm truthfully reports its low marginal cost of production. To ensure the firm finds it profitable to operate when its marginal cost is high, it must be the case that \( U_H \geq 0 \). Because social welfare declines as the firm’s equilibrium rent increases (when \( \alpha < 1 \) or \( \lambda > 0 \)), \( U_H \) is optimally held to zero. To limit the firm’s equilibrium rent, constraint (1) also will hold as an equality under the optimal regulatory policy, so \( U_L = \Delta q(p_H) \). Therefore, total expected welfare if price \( p_i \) is set when marginal cost \( c_i \) is realized (for \( i \equiv L, H \) ) will be:

\[ \frac{1}{2} [w_L(p_L) - [1 + \lambda - \alpha] \Delta q(p_H)] + \frac{1}{2} w_H(p_H) - [1 + \lambda] G, \] (2)

where \( G \) is a (fixed) cost of regulation that is financed with public funds. This cost might include the salary of the regulator and his staff, for example, and all other costs associated with acquiring essential information about the regulated industry.\(^{15}\) Differentiating expression (2) with respect to \( p_H \) reveals that when the regulator cannot observe the firm’s marginal cost, he is able to achieve the level of expected welfare that he could achieve if the firm’s costs were observable, but the high marginal cost was \( \hat{c}_H \), where:

\(^{15}\) \( G \) might also include the surplus lost when regulation retards industry innovation. Jackson et al. (1991) and Hausman (1997) estimate that delays in licensing wireless telecommunications providers in the United States reduced consumer surplus by billions of dollars.
\[ \hat{c}_H \equiv c_H + \left[ 1 - \frac{\alpha}{1 + \lambda} \right] \Delta > c_H. \tag{3} \]

Therefore, the optimal price when the high marginal cost is realized is the full-information (Ramsey) price corresponding to the inflated cost \( \hat{c}_H \). The inflated price when \( c_H \) is realized reduces the number of units of output on which the firm can exercise its cost advantage when \( c_L \) is realized. Therefore, the increase in \( p_H \) and the associated reduction in \( T_H \) limit the rent that accrues to the firm when its marginal cost is \( c_L \). Because the firm has no incentive to understate its production cost, there is no value to distorting the firm’s activities when it reports its marginal cost to be \( c_L \). Consequently, the optimal price when the low marginal cost is realized will be the full-information price corresponding to the firm’s actual cost \( (c_L) \).\(^{16}\)

Three features of the optimal regulatory policy in this simple setting characterize optimal regulatory policy in more general settings where the monopoly supplier is better informed than the regulator about key features of the regulated industry. First, the firm generally commands rent from its superior information. Second, to limit this rent and thereby generate greater surplus for consumers, the regulator will design a menu of options from which the firm can make a binding choice. A well-structured menu of options can induce the firm to employ its superior knowledge to secure outcomes that are better for both the firm and consumers in more favorable environments (e.g., when the firm has lower operating costs). Third, the optimal regulatory policy generally induces inefficient performance to limit the firm’s rent.

### 3.2 A Simple Model of Unregulated Competition

Now consider the following simple model of unregulated competition. Suppose two firms produce a homogeneous product and engage in Bertrand price competition. Each firm knows its own constant marginal cost of production and its rival’s marginal cost, \( c \in \{c_L, c_H\} \), when it sets its price. No transfer payments to or from the firms in the industry are possible in this setting. In particular, the firms’ profits cannot be appropriated by the government to reduce the general tax burden elsewhere. Therefore, the social cost of public funds \( \lambda \) plays no role in this setting, and social welfare is \( v(p) + \alpha \pi \), where \( \pi \) is industry profit. Both firms find it profitable to operate in

---

\(^{16}\) This fact is evident from differentiating the expression in (2) with respect to \( p_L \).
the industry.\footnote{Section 6 considers the possibility that both firms may not find it profitable to operate in the industry. See Auriol and Laffont (1992) and Riordan (1996), for example, for related models in which the duopoly is regulated.}

Each firm has the low marginal cost, \( c_L \), with probability \( 1/2 \). The firms’ costs may be correlated. Let \( \rho \in [1/2, 1] \) represent the probability that the two firms have the same cost.\footnote{The probability that both firms have high cost is \( \rho/2 \), the probability that both firms have low cost is \( \rho/2 \), and the probability that a given firm has low cost while its rival has high cost is \( 2(1-\rho) \).} The firms’ costs are perfectly correlated when \( \rho = 1 \). Their costs are uncorrelated when \( \rho = 1/2 \). Bertrand price competition ensures the equilibrium price will be \( c_H \) except when both firms have low cost, which occurs with probability \( \rho/2 \). A firm’s operating profit is zero unless it has the low marginal cost while its rival has the high marginal cost, in which case the firm’s profit is \( \Delta q(c_H) \).\footnote{The profit-maximizing price for a firm with the low marginal cost is assumed to exceed \( c_H \). Consequently, when only one firm has the low marginal cost, it will serve the entire market demand at price \( c_H \) in equilibrium.} A firm realizes this positive profit with probability \( [1-\rho]/2 \). Consequently, expected industry profit in this duopoly setting is \( [1-\rho]\Delta q(c_H) \), which declines as the firms’ costs become more highly correlated. Notice that the probability that the industry supplier has low marginal cost is \( (1-\rho/2) \). This probability decreases as \( \rho \) increases. In contrast, the probability that the industry price will be \( c_L \) is \( \rho/2 \), which increases as \( \rho \) increases.

Ignoring the firms’ fixed costs of production (\( F \)) for now, social welfare in this unregulated duopoly setting is:

\[
\frac{\rho}{2} v(c_L) + \left[ 1 - \frac{\rho}{2} \right] v(c_H) + \alpha \left[ 1 - \rho \right] \Delta q(c_H).
\]

(4)

3.3 Comparing Regulated Monopoly and Unregulated Competition.

Regulated monopoly offers four potential advantages over unregulated competition in this simple setting: (1) industry prices can be controlled directly; (2) transfer payments can be made to the firm to provide desired incentives; (3) the firm’s profit can be taxed to generate public funds, thereby reducing the deadweight losses associated with other sources of public funds; and (4) duplicative fixed costs of production can be avoided because there is only one industry supplier.

Unregulated competition has three corresponding potential advantages: (1) the likelihood that
the industry producer has the low marginal cost is higher than under monopoly because even if one
firm fails to secure the low cost, its rival may do so; (2) the presence of a rival with correlated costs
reduces the information advantage of the industry producer; and (3) any direct, operational costs of
regulation (e.g., the salaries of regulators and their staff) are avoided. The first of these potential
advantages of unregulated duopoly is referred to as the sampling benefit of competition. The second
potential advantage will be referred to as the rent-reducing benefit of competition.

To compare the performance of regulated monopoly and unregulated competition in this
setting, initially suppose \( \lambda = 0, G = 0, \) and \( F = 0, \) so there is no social cost of public funds, no
fixed cost of production, and no direct cost of regulation. Expressions (2) and (3) reveal that the
maximum expected welfare under monopoly regulation in this case is:

\[
\frac{1}{2} v(c_L) + \frac{1}{2} v(c_H + [1 - \alpha] \Delta). \tag{5}
\]

A comparison of expressions (4) and (5) provides four conclusions regarding the relative
performance of regulated monopoly and unregulated duopoly.

First, unregulated duopoly delivers a higher level of expected social welfare than does
regulated monopoly when the duopolists’ costs are perfectly correlated (so \( \rho = 1 \)). When costs
are perfectly correlated, the industry producer never has a cost advantage over its rival, and so
commands no rent in the duopoly setting. Furthermore, competition drives the industry price to the
level of realized marginal cost. Therefore, the ideal (full-information) outcome is achieved under
duopoly, but not under monopoly, where regulated prices diverge from marginal cost in order to
limit the rent the monopolist commands from its privileged knowledge of costs. In this case, then,
unregulated duopoly is preferred to regulated monopoly, even though the former offers no sampling
benefit. The benefits of competition arise entirely from rent-reduction in this case.

Second, when demand is perfectly inelastic, unregulated duopoly produces a higher level of
expected welfare than does regulated monopoly. When demand is perfectly inelastic, price
distortions do not affect output levels, and therefore do not serve to limit rent. Consequently, only
the probability of obtaining a low-cost supplier affects expected welfare, and this probability is
higher under duopoly than under monopoly because of the sampling benefit of competition.

\[\text{If } \alpha = 1, \text{ regulated monopoly and unregulated duopoly provide the same level of expected social welfare.}\]

\[\text{When demand is perfectly inelastic (so } q(\rho) \equiv 1, \text{ for example), expression (4) is weakly greater than expression (5) whenever } \rho / 2 \geq \alpha [\rho - \frac{1}{2}], \text{ which is always the case.}\]
Third, regulated monopoly will generate a higher level of expected welfare than unregulated duopoly when demand is sufficiently elastic (and $\rho < 1$). When demand is very elastic, prices that do not track costs closely entail substantial losses in surplus. Prices track costs more closely under regulated monopoly than under unregulated duopoly.22

Fourth, unregulated duopoly outperforms regulated monopoly when the difference between the high and the low marginal cost ($\Delta$) is sufficiently close to zero. Monopoly rent and duopoly profit are both negligible in this case, and so the choice between monopoly and duopoly depends upon which regime produces the low marginal cost more frequently. The sampling benefit of competition ensures the duopoly regime does so whenever costs are not perfectly correlated (so $\rho < 1$).23

More generally, if the fixed costs of operation ($F$) are sufficiently large, regulated monopoly will outperform unregulated duopoly in the simple model analyzed here because monopoly avoids the duplication of fixed costs.24 This conclusion is a corollary of the more general observation that monopoly supply will minimize industry costs if prevailing scale economies are pronounced relative to industry demand.25 Scale economies often are pronounced in network industries, where substantial physical infrastructure (e.g., a gas, water, or electricity distribution system or a telecommunications network) must be deployed in order to deliver service to customers.

When the social cost of funds ($\lambda$) is considered, regulated monopoly offers an additional advantage over unregulated duopoly. The monopolist’s rent can be taxed to fund desirable social projects, thereby reducing the need to employ other (potentially more costly) means to raise revenue.

22 The convexity of $v(\cdot)$ implies the expression in (5) is at least $[v(c_L) + v(c_H) - [1 - \alpha]\Delta q(c_H)]/2$. Therefore, the difference between expressions (5) and (4) is at least $[1 - \rho][v(c_L) - v(c_H)]/2 - \Delta q(c_H)(1 - \alpha)/2 + \alpha(1 - \rho)]$. Because this expression is non-decreasing in $\alpha$, it is at least $[1 - \rho][v(c_L) - v(c_H)]/2 - \Delta q(c_H)/2$. This term is positive when demand is sufficiently inelastic.

23 The benefits of unregulated competition may be less pronounced if industry competition is less pronounced. For example, Cournot competition may better describe industry interaction than Bertrand competition. Alternatively, industry producers might collude in setting prices. (Recall that limited industry competition may have contributed to the substantial increase in the wholesale price of electricity in California in 2000.)

24 In contrast, if the fixed cost of regulation, $G$, is sufficiently large, unregulated duopoly will outperform regulated monopoly because the former avoids this cost, by assumption.

25 In developing countries where consumers have limited income, the corresponding limited demand for key services can result in higher average production costs when scale economies are present.
4. Additional Considerations.

The simple models considered in section 3 abstracted from several important institutional factors that can affect the optimal choice between regulated monopoly and unregulated competition. These institutional factors include: (1) the resource constraints the regulator faces; (2) the potential role of regulation in pursuing distributional objectives; (3) the instruments available to the regulator; (4) the prevailing degree of regulatory independence and accountability; (5) the ownership structure of the incumbent industry producers; and (6) the importance of industry investment and innovation. These factors are now considered in turn.26

4.1 Resource Constraints.

Although the regulator was not omniscient in the models of section 3, he had considerable knowledge of the regulated industry. In practice, a regulator’s information can be far more limited.27 A regulator’s difficult task of overseeing and directing the activities of a monopoly supplier can become nearly impossible when the regulator’s information and expertise are severely limited and when he lacks the physical and financial resources to overcome these limitations. Consequently, allowing competition to replace regulatory oversight as the primary means of motivating and disciplining the incumbent supplier can be advantageous when the efficacy of regulatory oversight is severely compromised by limited regulatory resources.28

The resources of other oversight agencies also influence the relative merits of monopoly supply and liberalization. For example, competition may be less likely to impose meaningful discipline on a dominant incumbent supplier if the antitrust agency that oversees industry competition has limited expertise and meager physical and financial resources. Consequently, both the relative and absolute level of resources available to oversight agencies influence the relative merits of monopoly and

26 For further discussion of the effects of institutions on regulatory policy, see Levy and Spiller (1994, 1996), Newbery (1999), Noll (2003), Guasch (2004), Kessides (2004), Jamison et al. (2005), and Laffont (2005), among others. We focus on the effects, rather than the origins, of a country’s institutions. Acemoglu et al. (2005) trace a country’s institutions to its colonial origins. In countries where poor living conditions rendered settlement unattractive, for example, little effort was devoted to developing the country’s institutions. Such effort was more prominent in countries where long-term settlement was more attractive and more widespread.

27 The theory of optimal regulation when the firm is privately informed about several aspects of its operation awaits further development. Rochet and Stole (2003) provide a survey of the theory of multidimensional screening.

28 Stern (2000) documents the limited regulatory resources that are available in many countries. Stern suggests that resource sharing among regulatory agencies can help to mitigate partially the effects of severe shortages of critical regulatory resources in some settings. Noll (2003) also emphasizes the merits of sharing regulatory resources.
4.2 Income Redistribution and Universal Service Goals.

The analysis in section 3 abstracted from any differences in the cost of serving different customers and did not model explicitly differences in the wealth levels of different consumers. Such differences can raise concerns about universal service – the ubiquitous delivery of essential services at affordable rates. When a government has limited ability to redistribute income directly (perhaps because income taxes are widely evaded, for example), the regulated prices of essential services can constitute an important means to promote universal service. To illustrate, suppose a country wishes to ensure that all its citizens have access to clean water or to basic communications services at low prices. Because prices tend to reflect costs under unfettered competition, individuals that are particularly costly to serve (because of their geographic location, for example) may face unduly high prices for key services under unregulated competition. In contrast, a regulated monopolist will agree to serve high-cost (e.g., rural) customers at relatively low prices if it is permitted to offset the associated financial losses by charging prices sufficiently above the costs of serving low-cost (e.g., urban) customers. Consequently, universal service (and other distributional) concerns can cause regulated monopoly to be preferred to unregulated competition.

4.3 Available Instruments.

The policy instruments available to oversight agencies also influence the relative merits of monopoly and competition. To illustrate, return to the simple setting considered in section 3, where the regulator was able to set prices and deliver transfer payments. In practice, regulators are not always able to deliver transfers to the firms they regulate. If the regulator lacked this ability in the setting analyzed in section 3.1 and wished to ensure the monopolist never terminated its operations, the regulator could do no better than to set a single price equal to the high marginal cost, \( p = c_H \) (assuming fixed costs are zero). This policy would generate expected welfare

\[
v(c_H) + \frac{1}{2} \alpha \Lambda q(c_H).
\]

It is apparent that the level of expected welfare in expression (6) is less than the corresponding level

---

29 The public finance literature (e.g., Atkinson and Stiglitz, 1976) notes that when a nonlinear income tax is feasible, prices of goods and services often should reflect production costs. Laffont and Tirole (2000, section 6.2) consider the implications of this conclusion for pricing policy in regulated industries. Burgess and Stern (1993) review the theory and the practice of tax policy in countries with limited wealth.
in expression (4). Consequently, unregulated duopoly is always preferred to this restricted form of monopoly regulation in the setting considered in section 3.1.

More generally, if a regulator has limited ability to reward a monopoly supplier for superior performance and penalize the firm for inferior performance, the regulator may be unable to motivate the firm to serve consumers well. This is the case regardless of how well the regulator understands the firm’s capabilities and consumers’ preferences. Similarly, if the regulator is not authorized to compel the firm to report data on its operations, the regulator will find it difficult, if not impossible, to make informed policy decisions, even if the regulator has ample resources and ability to analyze and interpret data. Consequently, limited powers to reward or penalize the firm or to compel data reporting can render monopoly regulation ineffective, and so can increase the relative merits of unregulated competition.

A regulator’s powers to control the activities of new suppliers also can affect the relative merits of monopoly and competition. To illustrate, suppose the regulator cannot impose any regulations on new competitors. In particular, suppose a regulator cannot limit the range of services that competitors offer or tax any of these services. In this case, even though liberalization may help to motivate the incumbent supplier to reduce its operating costs, it may allow competitors to engage in cream-skimming. Cream-skimming is the act of serving the most profitable (e.g., urban, business telecommunications) customers, and leaving the incumbent supplier to serve the less profitable (e.g., rural, residential telecommunications) customers. Cream-skimming can limit the ability of the incumbent supplier to finance particularly low prices on some services with substantial profit earned on other services, and thereby undermine socially desirable pricing structures. Therefore, even though managed competition (which entails the regulation of both incumbent suppliers and new entrants) might be preferable to monopoly, monopoly may be preferable to unfettered competition when an ineffective tax system requires that universal service be pursued through industry prices.

---

30 The difference between expressions (4) and (6) is $[\nu(c_L) - \nu(c_H)] \rho/2 + \alpha \Delta q(c_H)\left[\frac{1}{2} - \rho\right]$. The convexity of $\nu(\cdot)$ implies this expression is at least $[\rho/2 - \alpha(\rho - \frac{1}{2})] \Delta q(c_H)$. This expression is non-negative because it is a decreasing function of $\rho$ and is non-negative at $\rho = 1$.

31 See Curien et al. (1998) and Laffont and Tirole (1990a), for example.

32 Of course, a country with limited ability to collect taxes also may have limited ability to enforce desired prices. The relative strengths of the country’s institutions are important to consider in such cases. A country may be better able to measure electricity consumption and enforce associated charges than to prevent citizens from hiding wealth in order to evade income taxes, for example.
4.4 Private vs. State Ownership.

The extent of government ownership of the dominant incumbent supplier also can affect the both the merits and the most appropriate form of liberalization. A firm that is largely owned by the government can be less responsive to the oversight and demands of shareholders than are privately-owned firms. Furthermore, state-owned enterprises (SOEs) can face softer budget constraints than their privately-owned counterparts, in the sense that the government may be more willing to tolerate losses by the firm and to finance the firm’s continued operation despite poor historic financial performance. Because of the corresponding diminished incentive to minimize production costs, an SOE may operate with higher costs than a privately-owned monopoly. Therefore, liberalization may have greater potential to reduce industry costs when the incumbent monopoly supplier is owned primarily by the government than when it is owned primarily by private investors.

4.5 Regulatory Independence and Accountability.

The simple models considered in section 3 presumed the regulator faithfully pursued the social objective. In practice, a “captured” regulator may not do so. A regulator is said to be captured by the firm he regulates when the regulator generally implements policies that further the interests of the firm at the expense of the broader social interest. Many factors increase the likelihood of regulatory capture. For example, a regulatory agency with limited expertise and resources may be forced to rely heavily on the advice and information supplied by the firm when formulating policy. Alternatively, the firm may routinely offer attractive employment opportunities to regulators who have proved to be cooperative, and the country’s laws may not preclude such offers. Also, the firm may provide a sizable portion of the regulator’s ongoing budget, and may have some discretion over the timing and magnitude of its contributions to the regulator’s budget. When factors like these lead

---

33 For instance, Cragg and Dyck (2003) find that the financial compensation of managers is not as closely linked to the performance of the firm in state-owned enterprises as it is in privately-owned firms.

34 See Kornai et al. (2003).

35 Megginson et al. (1994), D’Souza and Megginson (1999), La Porta and López-de-Silanes (1999), Dewenter and Malatesta (2001), Megginson and Netter (2001), Wallsten (2001), and Djankov and Murrell (2002), among others, present (sometimes mixed) evidence that the efficiency of SOEs increases after privatization, particularly in the presence of substantial industry competition and independent regulators.

36 Of course, in practice, this greater potential need not always translate into more pronounced reductions in industry costs. Mizutani and Uranishi (2003), for example, report that liberalization in Japan’s postal industry did not substantially increase the productivity of the SOE.


38 See Che (1995) and Salant (1995), for example.
to a high likelihood of regulatory capture, the entry of additional competitors may be the best way to impose meaningful discipline on the incumbent supplier and otherwise ensure long-term gains for consumers.\footnote{Holmstrom and Milgrom (1991), Holmstrom (1999), and Dewatripont et al. (1999) examine the effects of long-term career concerns on short-term behavior. Such concerns can, for example, induce a regulator to focus primarily on performance dimensions that are readily observed (e.g., short-term price reductions) and less on performance dimensions that are more difficult for potential employers to observe in a timely fashion (e.g., prospects for viable long-term industry competition).}

The importance of regulatory independence from short-term popular opinion also is apparent. Producers in network industries typically incur substantial sunk (non-recoverable) costs. These producers also often deliver services to a large portion of the population, and so the prices of these services are of substantial public concern. Together, these two elements create substantial risk of expropriation by well-meaning, but short-sighted regulators. In response to public pressure, regulators may reduce prices as far as possible toward variable production costs. As long as regulated prices allow the firm to recover its variable production costs, the firm will prefer to continue to produce than to terminate its operations. Thus, in the short run, the regulator gains by securing low prices and ongoing production of key services. Although such a policy may provide short-term gains, it can have substantial long-term costs. Potential and actual producers will realize they are unlikely to recover any sunk costs they incur. Consequently, they will be reluctant to incur such costs, and so existing network infrastructure will be permitted to decay and new network infrastructure will not be built. Thus, if regulators are to design and implement policies that best serve the long-term interests of consumers, they must be able to develop policy credibility by resisting short-term pressures to renege on long-term promises.\footnote{See Williamson (1975) for a pioneering treatment of the problem, and Newbery (1999) for an extensive discussion of the problem of regulatory commitment. Some authors (e.g., Henisz, 2000) have developed indices to measure expropriation risk in different countries. Jamison et al. (2005) review these indices.}

A regulator’s commitment powers can be enhanced by a variety of factors, including strong legal institutions and a long tenure. Strong legal institutions can thwart attempts by other government agencies to intervene in the day-to-day operations of the regulatory agency, and can thereby enhance a regulator’s commitment powers by reducing the likelihood that the terms of announced regulatory policies will be changed.\footnote{Wallsten (2001) finds that the privatization of state-owned telecommunications providers in Latin America and Africa is associated with improved industry performance when the industry regulator is independent (in the sense of not being directly under the control of a government ministry), but not otherwise. Edwards and Waverman (2005) find that regulatory independence is associated with lower charges for access to the network infrastructure of state-owned incumbent suppliers of telecommunications services.} In particular, strong legal institutions can enforce
long-term contracts between the regulator and the firm, and can prevent a regulator from changing the terms of announced policies (including promised returns on investment) in response to pressure from other government agencies or the citizenry at large. When legal institutions are weak, pressure groups may anticipate substantial benefits from convincing the regulator to renege on the promises he has made to the firm. Thus, the independence of a country’s judicial system and the ability of a country to enforce the terms of legal contracts can affect the optimal design of industry policy.

A regulator may also be better able to pursue policies (such as delivering promised rewards to the regulated firm) that promote the long-term, rather than the short-term, interests of consumers if his tenure as regulator is relatively long. Long-lived regulators with sufficient concern about future industry outcomes will realize that short-term expropriation of an incumbent producer’s investment will discourage future investment by the same firm or its successor. In addition, a regulator may feel less pressure to pander to popular opinion if he is not elected by direct vote of the citizenry. Thus, a setting where the regulator serves a fairly long term and faces reappointment by the government may provide an appropriate trade-off between independence and accountability.

There is an additional benefit of a reasonable degree of regulatory independence from direct intervention by other government agencies. A government that is unable to intervene in the short-term operations of a regulatory agency is compelled to state as fully and clearly as possible the mission of the regulatory agency. Failure to do so will afford the regulatory agency the opportunity to pursue its own mission and goals, rather than those of the government more broadly. A coherent and transparent statement of the principles that will guide industry policy provides greater certainty for industry participants, which can encourage investment and facilitate long-term

42 Kessides (2004, p. 102) describes settings where governments have changed the terms of legal contracts.

43 See Newbery (1999), Salant and Woroch (1992), and Levine et al. (2005) for formal analyses of this effect. Levine et al. (2005) also emphasize the potential value of delegating authority to an independent regulator with substantial concern for the welfare of the firms he regulates.

44 Besley and Coate (2003) employ data from the United States to examine the effects of the method by which state regulators are selected on performance in the state’s electric power industry. The authors find lower prices and reduced investment (as proxied by the frequency of power outages) in states where regulators are elected than in states where regulators are appointed.

45 See Maskin and Tirole (2004) for a formal analysis of this trade-off.

46 In this respect, regulatory independence can deliver benefits similar to those provided by the privatization of a state-owned enterprise. Privatization increases the cost to the government of intervening in the day-to-day operations of the firm, and thereby renders more credible the government’s promise to refrain from such intervention (Sappington and Stiglitz, 1987). Regulatory independence enhances the commitment powers of regulators in similar fashion, once the mission and goals of the regulatory agency are stated clearly.
When regulatory capture is likely, explicit restrictions on the regulator’s autonomy and commitment powers may be desirable in order to limit the regulator’s ability to pursue interests other than the long-term social interest. For example, key powers might be dispersed among multiple regulators or divided between a regulator and other government agencies. Alternatively or in addition, the regulator’s discretion in formulating policy might be substantially restricted. The tenure of individual regulators might also be limited.

Competition can play a particularly valuable role in disciplining and motivating the incumbent supplier to pursue the best interests of consumers when the regulator’s powers to do so are (inherently or intentionally) limited. However, competition is not necessarily a panacea. Weak commitment powers may limit non-fungible investment under both unregulated competition and regulated monopoly. Competitors, like the incumbent supplier, will recognize that their investments may be expropriated by a regulator with substantial expropriation powers and limited commitment powers, and so may be reluctant to undertake the investment required to improve industry performance. Thus, there is seldom an effective substitute for strong commitment powers.

---

47 Spiller and Vogelsang (1997) and Bakovic et al. (2003) emphasize the value of coupling substantial regulatory independence with a clearly specified regulatory contract (e.g., a regulatory license). Under such coupling, regulators will have the independence required to implement impartially the details of a politically popular contract, but will not be empowered to implement any policy of their choosing. Using a large dataset of monopoly franchise auctions, Guasch (2004, Table 1.16) reports that the original terms of the franchise contract were renegotiated 61% of the time when there was no separate regulatory body responsible for contract administration. The corresponding percentage was only 17% when the contracts were administered by a separate regulatory body.

48 The independence granted the telecommunications regulator in Jamaica is believed to have led to a decline in industry investment, as investors feared the regulator would use his autonomy to expropriate investments (Levy and Spiller, 1994).

49 See Laffont and Martimort (1999) and Martimort (1999b).

50 See Laffont and Tirole (1990b).

51 Martimort (1999a) shows that it can be optimal to impose more severe restrictions on a regulatory agency over time, as the potential for regulatory capture increases.

52 If competition undermines a regulator’s (otherwise substantial) powers to promise attractive returns to the incumbent monopoly supplier, it is possible that competition may be inadvisable when the incumbent monopolist is investing and otherwise satisfying the long-term needs of consumers reasonably well. In extreme cases, unregulated monopoly might be the most effective way to deliver strong investment incentives if the absence of regulation, itself, constitutes a meaningful commitment not to expropriate a monopolist’s earnings. Notice that although a regulator may have particularly pronounced ability to expropriate investors because of his direct control over a broad range of industry policies, other government entities also may have substantial ability to expropriate investors, by imposing excessive profit taxes or indirect taxes, for example.
4.6. Investment and Innovation.

The link between industry structure and the incentive to innovate and to reduce costs is complex, even in the absence of regulation. Competing firms may have greater incentive than an unregulated monopoly provider to reduce operating costs in part because industry output (and thus the potential cost savings from a reduction in marginal production cost) is greater under competition than under unregulated monopoly. \(^5\) In contrast, substantial market concentration can encourage innovation for at least two reasons. First, the profit a monopolist generates can serve as a valuable source of research and development (R&D) funding. \(^5\) Second, the prospect of substantial monopoly profit can be a compelling reason to undertake R&D investment. \(^5\)

Regulatory policy can affect infrastructure investment differently than it affects innovative effort and investment designed to reduce operating costs. \(^5\) To illustrate this point, first consider rate of return regulation, which promises a fair return on prudently incurred investment. When expropriation can be avoided, such a promise can deliver strong incentives for infrastructure investment. In contrast, because it requires revenues to track costs closely, rate of return regulation (like other forms of “cost-plus” regulation) typically provides limited incentive for innovation and cost reduction. \(^5\) Now consider price cap regulation, which typically permits revenues to diverge from realized costs for a specified period of time (e.g., four years), but does not promise specific long-term returns on investment. Although such a policy can provide substantial incentive for short-term innovation and cost reduction, \(^5\) it may provide limited incentive for long-term infrastructure investment.

\(^5\) Cohen and Levin (1989) review empirical studies of the relationship between market structure and innovation.

\(^5\) See Arrow (1962). Of course, this argument presumes that intellectual property protection precludes competitors from immediately copying the innovator’s discovery, and thereby eliminating the financial gain from innovation.

\(^5\) This argument presumes that capital markets are imperfect. This is a reasonable assumption in the context of innovation because innovators often are unable to convince investors of the merits of their potential innovation without revealing the details of the innovation (and thereby forfeiting some of the potential financial gains from the innovation). See, for example, Anton and Yao (1994, 2002).

\(^5\) See Schumpeter (1950) and Loury (1979), for example.

\(^5\) Guthrie (2006) provides a more complete review of the literature that examines the impact of regulation on infrastructure investment.

\(^5\) A regulatory policy that delivers no extra profit to a firm as its realized production costs decline effectively expropriates any investment the firm might make in an attempt to secure lower production costs.

\(^5\) The regulatory policy implemented for British Gas in 1986 (which permitted the average price of natural gas to increase at the rate $RPI+Y-X$) is an example of price cap regulation. Sappington (2002) provides an overview of price cap regulation plans. Also see Acton and Vogelsang (1989), Hillman and Braeutigam (1989), Armstrong et al. (1994), and Bernstein and Sappington (1999), for example. Notice that a promise...
Therefore, the choice between rate of return regulation and price cap regulation will depend in part on the type of investment that is most important to secure. In settings where the top priority is to induce the regulated firm to employ its existing infrastructure more efficiently, price cap regulation may be preferable. In contrast, in settings where it is important to reverse a history of chronic under-investment in key infrastructure, rate of return regulation may be preferable.

The appropriate choice between rate of return regulation and price cap regulation also is influenced by industry volatility and regulatory commitment powers. As costs and demands change over an extended time period, prices and costs will invariably diverge under price cap regulation, possibly leading either to financial distress or to particularly large profit for the regulated firm. Neither of these outcomes is likely to be credibly sustained in settings where the country’s institutions are weak. Therefore, in such settings, rate of return regulation may be preferable to price cap regulation in the presence of considerable industry volatility, particularly if infrastructure investment is desirable. The cost-plus nature of rate of return regulation, which ensures profits are neither excessive nor insufficient, can render its implicit commitment to set prices that track costs closely more credible than the price commitments encompassed in a price cap plan.

5. Franchise Bidding.

Even when pronounced scale economies render direct competition in the market prohibitively to allow revenues to diverge from realized costs can play much the same role as intellectual property rights protection that allows an innovator to derive substantial financial gain from her innovation.

Spence (1975) observes that when quality cannot be regulated directly, rate of return regulation may provide stronger incentives to supply quality than price cap regulation. Sappington (2005a) reviews the literature on service quality regulation in utility industries.

Regulatory policies that deliver a consistent rate of return on investment also can reduce the variance in investors’ financial returns, and thereby reduce the regulated firm’s cost of capital.

Alternatives to these two forms of regulation also merit consideration. As Schmalensee (1989) notes, earnings sharing plans (which specify explicit rules for sharing the firm’s realized earnings with consumers) can help to avoid extreme distributions of rent that can prove difficult to enforce in practice. Sappington (2002) provides a discussion of earnings sharing plans and relevant references to the literature. Vogelsang (2005) explains how earnings sharing policies might be coupled with other regulatory regimes to provide appropriate short-term incentives for cost reduction and long-term incentives for investment.

This is the case even if the price cap plan makes explicit adjustments for changes in economy-wide output prices and changes in key input prices.

Of course, if a regulator has considerable discretion over which of the firm’s investments qualify for the promised rate of return, serious problems of opportunism can arise under rate of return regulation, just as they can arise under price cap regulation.

Guasch (2004, Table 1.16) reports more widespread renegotiation of franchise contracts under price cap regulation than under rate of return regulation.
costly, competition for the right to serve as a regulated monopoly supplier (in the form of franchise bidding) can sometimes capture for consumers much of the surplus that strong competition in the market would generate. To illustrate the potential value of franchise bidding, return to the simple model considered in section 3 and suppose two firms bid for the right to serve as the sole provider of the product in question under terms specified by the regulator. Each firm is privately informed about its own marginal cost of production \( c \in \{ c_L, c_H \} \). The two possible cost realizations again are equally likely for each firm. Furthermore, for simplicity, the firms’ cost realizations are uncorrelated.

The optimal franchise bidding policy in this setting takes the following form. The two firms report their cost realizations simultaneously. If only one firm reports the low cost realization \( c_L \), that firm is selected to be the monopoly supplier, sell the product at unit price \( p_L \), and receive transfer payment \( T_L \). If both firms report low costs, one of the firms is selected at random to operate the \((p_L, T_L)\) contract. If both firms report high costs \( c_H \), one of the firms is selected at random to operate under the \((p_H, T_H)\) contract.

The contracts are optimally designed to ensure the firms report their costs truthfully. Consequently, a firm with the low marginal cost will be selected to serve as the monopoly supplier with probability \( \frac{3}{4} \). (Thus, franchise bidding secures the sampling benefit of competition.) A firm with the high marginal cost will operate with probability \( \frac{1}{4} \). Truthful reporting by the low cost firm will be a best response to truthful reporting by the rival potential producer if

\[
\frac{3}{4} U_L \geq \frac{1}{4} \left[ U_H + \Delta q(p_H) \right],
\]

where \( U_i = [p_i - c_i] q(p_i) - F + T_i \) for \( i = L, H \). The binding participation constraint will be that of the firm with the high marginal cost, so \( U_H \) is optimally held to zero. Inequality (7) will optimally be satisfied as an equality, to limit the rent afforded a firm with the low marginal cost. Consequently, expected welfare under the optimal pricing policy is:

\[
\frac{3}{4} \left[ w_L(p_L) - \frac{1}{3} \left[ 1 + \lambda - \alpha \right] \Delta q(p_H) \right] + \frac{1}{4} w_H(p_H).
\]

Demsetz (1968) provides the seminal discussion of this observation. Laffont and Tirole (1987), McAfee and McMillan (1987), and Riordan and Sappington (1987), among others, provide formal analyses of the optimal design of monopoly franchises.
Expression (8) reveals that $p_H$ is optimally chosen to maximize $w_H(\cdot) - [1 + \lambda - \alpha] \Delta q(\cdot)$ while $p_L$ is optimally chosen to maximize $w_L(\cdot)$. Therefore, prices in this franchise bidding setting are the same prices that are implemented in the regulated monopoly setting. (Recall expression (2).) The transfer payment ($T_L$) to the low-cost supplier is reduced in the franchise bidding setting, though. The firm will reveal its superior capabilities despite being promised a smaller transfer payment because cost exaggeration entails substantial risk of being excluded from the industry. (Thus, franchise bidding secure the rent-reducing benefit of competition.)

Because franchise bidding and monopoly regulation implement the same price for a given cost realization, total expected welfare is the same under the two regimes except that the likelihood of a low-cost supplier increases from $1/2$ to $3/4$. In the case where there is no cost of social funds (so $\lambda = 0$), expected welfare in the franchise bidding setting is:

$$\frac{3}{4} v(c_L) + \frac{1}{4} v(c_H + [1 - \alpha] \Delta).$$

(9)

It is apparent that expression (9) exceeds expression (5). Expression (9) also exceeds expression (4) when $\rho = 1/2$. \(^{68}\)

In summary, franchise bidding outperforms both monopoly regulation and duopoly competition in this simple setting. \(^{69}\) It does so by employing both transfer payments and price regulation to pursue social goals, by ensuring the benefit of scale economies through the selection of a single supplier, and by securing the sampling and rent-reducing benefits of competition.

More generally, franchise bidding can have its drawbacks. If potential operators do not have comparable skills, accurate information, and substantial financial resources, bidding for the right to

\(^{67}\) The regulator can achieve the ideal (full-information) outcome in this setting if the firms’ costs are correlated. See, for example, Demski and Sappington (1984) and Crémer and McLean (1985).

\(^{68}\) This is the case because $v(c_H + [1 - \alpha] \Delta) \geq v(c_H) - [1 - \alpha] \Delta q(c_H)$. Therefore, expression (9) will exceed expression (4) if $2[v(c_L) - v(c_H)] \geq \Delta q(c_H) [1 + \alpha]$. This inequality holds for all values of $\alpha \leq 1$.

\(^{69}\) In practice, franchise auctions often take different forms. For example, firms often are invited to specify a price at which they are willing to supply the product in question to consumers, and the firm that bids the lowest price might be awarded exclusive production rights. This auction, which involves no transfer payments from firms to the government (or vice versa), typically will result in the most efficient firm serving consumers at a price that reflects its average cost of production. Such an auction resembles normal price competition, but has the added advantage that it can permit profitable operation even in the presence of scale economies. Alternatively, the supply price might be determined in advance, and firms could bid on the amount they will pay the government (or the payment they will require from the government) for the right (and obligation) to provide the service at the stipulated price.
serve as the monopoly supplier may not be intense. Firms will have little interest in bidding for a contract they expect to lose to a more capable operator. In addition, a firm will be reluctant to bid aggressively on a contract if it may win the contract not because it is the most capable producer but because it underestimates most severely the true (common) cost of operating the franchise. A firm may be similarly reluctant to bid for a contract when it suspects other bidders (e.g., an incumbent provider of the service in question or a firm that provides similar services in another geographic region) have better information regarding the financial returns the contract is likely to provide. Furthermore, a firm cannot bid more for a contract than its financial resources allow. When pronounced differences in skills or information and/or a dearth of qualified bidders with substantial financial resources limit the intensity of franchise bidding, such bidding will not capture for consumers the surplus they would enjoy if multiple, well-informed firms with similar capabilities competed against each other in the marketplace.

Franchise bidding also may provide opportunities for the selected producer to “hold up” the government. After it is awarded the monopoly franchise, the chosen producer may attempt to renegotiate the terms of the original contract in order to secure more favorable terms. The government may be susceptible to renegotiation demands in order to avoid the appearance of failure in the procurement process or to avoid substantial transaction costs associated with re-auctioning the franchise.

---

70 This possibility is known as the “winner’s curse”. See Krishna (2002, pp. 84-85), for example.

71 The terms of the franchise contract can be altered to counteract the problems that arise when bidders have limited financial resources. For example, rather than bid on a lump-sum franchise fee, potential operators might bid on the fraction of realized profit (or revenue) that they are willing to share with their customers. See Lewis and Sappington (2000) for details. Che and Gale (1998, 2000) provide related analyses.

72 A decision to exclude foreign investors from the bidding process (to promote nationalism or to further national security, for example) can limit substantially the intensity of franchise bidding. Such exclusion also can reduce the value of the franchise, since the skills and expertise of the successful bidder affect the profit the enterprise can generate. Frydman et al. (1999) report that enterprise revenues increase substantially when an SOE is sold to outside managers, but not when it is sold to individuals who managed the SOE before its privatization.

73 Guasch (2004) analyzes more than 1,000 franchise auctions in Latin America and the Caribbean between 1985 and 2000. He finds that more than 50% of electricity franchise contracts and 75% of water franchise contracts were renegotiated. The average time between franchise award and renegotiation was approximately two years. Renegotiation was initiated by the chosen operator more often than it was initiated by the government.

74 When all relevant dimensions of performance are not specified clearly in a regulatory contract, a potential supplier may be able to bid the highest franchise fee, not because it is the least-cost supplier of the services in question, but because it will deliver the least on all of the performance dimensions that are not specified in the regulatory contract. Manelli and Vincent (1995) identify conditions under which considerations of this sort render it optimal for a regulator to negotiate a contract with a single selected
Franchise bidding also can fail to provide ideal incentives for investment. If the duration of the franchise contract is short relative to the useful life of a desirable sunk investment, the chosen supplier may be reluctant to undertake the investment if the firm’s tenure as the monopoly supplier is likely to end before it can fully recover the cost of the investment. Consequently, the chosen supplier may adopt an inefficient production technology (one that employs an unduly small level of sunk costs), and all potential suppliers may reduce their bids for the right to serve as the monopolist.

Although long-term contracts can, in principle, help to overcome this problem, long-term contracts seldom are a panacea. In practice, it typically is impossible to delineate all relevant contingencies in a contract. Furthermore, rigid contracts can preclude valuable adaptations to changing industry conditions. In addition, even the most carefully crafted long-term contracts may not be enforced if the prevailing legal institutions are weak. For all these reasons, even long-term contracts typically are unable to deliver ideal investment incentives. In some settings, it may be possible to adjust franchise bidding policies to counteract some of these problems. For example, auction rules that favor the incumbent supplier can enhance the incumbent’s incentive to undertake sunk investments in the presence of short-term contracts.

In summary, auctions for the right to be the sole supplier can help to limit monopoly rents and to select the most efficient industry supplier. However, franchise auctions seldom eliminate the need for regulation. Moreover, the same strong regulatory institutions that are necessary for effective monopoly regulation are required for the success of franchise bidding contracts. When these institutions are present, though, franchise bidding can constitute a useful additional instrument that regulators can employ to select and discipline sole providers in network industries.

In concluding this section, we note that yardstick competition can secure many of the same benefits as franchise bidding in settings where different monopolists operate in distinct geographic

---

75 See Williamson (1976), for example.
76 Prager (1989), Zupan (1989a,b), and Otsuka (1997) present evidence which suggests that these difficulties with franchise competition are not always insurmountable in practice.
77 See Laffont and Tirole (1988). Investment also might be encouraged by adopting technologies that employ more fungible physical assets (e.g., wireless rather than wireline facilities in the telecommunications industry).
78 Regulators can sometimes benefit by ensuring that alternative suppliers are available to replace the incumbent supplier, as needed. Although it can be costly to maintain a second source of production, the ability to readily shift some or all production to the second source can impose useful discipline on a monopoly supplier. See, for example, Rob (1986), Anton and Yao (1987), Demski et al. (1987), Riordan (1996), Riordan and Sappington (1989), Dana and Spier (1994), and Anton and Gertler (2004).
markets. (For example, different water distribution companies often serve different regions of a country.) In such settings, each monopolist might be compensated on the basis of how its performance compares to the performance of other monopolists. For example, the compensation delivered to each firm might be set equal to an index of the realized costs of the other firms, rather than its own costs. This and other forms of yardstick competition can provide strong incentives for efficient performance by all monopolists when they are known to operate in similar settings. When the firms operate in environments that differ substantially (in geographic area, terrain, weather, or population density, for example), explicit corrections for relevant differences can be important. Such corrections typically will be necessary to avoid compensation that is unduly generous for some firms and unduly meager for others. Appropriate handicapping can be difficult in the presence of limited information about the precise nature of the variation in the firms’ operating conditions. However, some relative performance comparisons generally can help to discipline and motivate monopoly suppliers.79


The foregoing discussion suggests the greatest potential gains from competition will tend to arise when: (i) industry scale economies are limited relative to consumer demand; (ii) the industry regulator has limited information, limited resources, and limited instruments with which to craft policy; (iii) the regulator’s commitment powers are limited; and (iv) subsidization of the consumption of some of the dominant supplier’s services is either not critical or can be achieved by means other than regulating the supplier’s price structure.

The design of liberalization policy is of paramount importance in settings where competition is deemed to be a superior alternative to a prevailing monopoly regime. The purpose of this third segment of the present essay is to discuss the principles that underlie the design of sound liberalization policy.80 After considering the subtle issue of explicit entry assistance, this section focuses on liberalization policies that can hinder (rather than promote) vibrant, long-term industry competition. Section 7 reviews preferable competition-enhancing liberalization policies.

It is important to emphasize at the outset that the discussion in both this section and the next

79 See Shleifer (1985), Sobel (1999), Carrington et al. (2002), and Farsi et al. (2005), for example. Armstrong and Sappington (2005, section 4.1) provide further discussion and references. Notice that the entire benefit of yardstick competition arises from the rent-reducing impact of operation by multiple firms. Yardstick competition does not admit a sampling benefit because the firm that supplies a given market is fixed exogenously.

80 This discussion draws in part from Armstrong et al. (1994, chapter 4).
necessarily entails some subjective judgments. Furthermore, even though the policies considered in this section generally are not recommended, some of the policies can, in theory, enhance welfare in certain settings if the regulator is particularly well informed. Therefore, definitive, unequivocal conclusions about liberalization policies are difficult to draw. The ensuing discussion is intended to provide general guiding principles (to the extent possible) rather than precise, comprehensive prescriptions and definitive conclusions.

Successful liberalization is seldom as simple as removing all legal restrictions on entry into the regulated industry. Entrants face myriad economic barriers to entry, even when legal barriers are removed. These entry barriers include: (1) customer inertia due to switching costs or ignorance, for example; (2) incumbent control of key inputs that entrants require for profitable operation; and (3) the prospect of aggressive pricing by incumbent suppliers. Policies that reduce or limit the effects of entry barriers constitute vital components of a successful liberalization plan, as explained in section 7. However, is the reduction of entry barriers sufficient, or should liberalization efforts include additional policies that provide explicit assistance to competitors?

Although such direct assistance has many important drawbacks (as explained further below), it can in principle increase total welfare under some conditions. To illustrate, suppose competition would drive prices close to the entrant’s marginal cost of production, as it would, for example, in the setting of section 3.2 when the competitors’ costs are highly correlated. In such a setting, an entrant will anticipate limited variable profit from industry operations. Consequently, if the entrant must incur a substantial sunk cost in order to enter the industry, the potential competitor will decline to enter, even if it faces no legal restrictions on entry. Therefore, entry may be least likely to occur without assistance precisely when, due to its intensity, competition is most effective relative to regulation.81 In such settings, entry assistance (perhaps in the form of a subsidy financed by public funds to cover the sunk entry cost, for example) could increase social welfare in theory. However, the appropriate magnitude of such entry assistance can be difficult to determine in practice. Furthermore, such assistance can promote regulatory capture and unproductive use of public funds.

The general merits of entry assistance depend in part on the external effects of entry. External effects refer to the effects of entry on consumer surplus and the profits of incumbent producers. External effects may cause entry to be unprofitable for an individual firm even though entry would increase consumer surplus and total welfare, as in the setting just described. External effects also

\[81\text{Entrants may require little assistance when they are competing against an inefficient incumbent supplier that has historically delivered poor service quality to consumers. Consumers often will be anxious to secure service from an alternative supplier under such conditions.}\]
may cause entry to be profitable for an individual firm even though entry would reduce total welfare. Such a setting (in which it can be more appropriate to discourage entry than to assist entry) arises naturally when the profit an entrant anticipates from industry participation is derived in part from profit that would otherwise accrue to incumbent producers.82

Because the most appropriate level of direct entry assistance can be difficult to determine and because such entry assistance can have undesirable consequences,83 it generally is not recommended. The following liberalization policies (some of which entail entry assistance and some of which fail to reduce entry barriers adequately) also generally are not recommended.

6.1 Provide a Temporary Monopoly or Oligopoly.

A common form of liberalization is to announce a future (sometimes distant) date at which competition will be admitted, but explicitly preclude widespread competition before that date.84 A temporary monopoly (or oligopoly) policy of this sort often is adopted in part to secure the incumbent supplier’s support for liberalization. Such support may be important, for example, when the incumbent supplier is a state-owned enterprise with substantial authority to set industry policy or when the incumbent supplier employs a large labor force and so has substantial political power.

In principle, a temporary monopoly policy can have some merit. Intense competition can increase the risk of and limit the return from investment, and thereby reduce the investment of the incumbent supplier. Therefore, in principle, a temporary monopoly could increase the incumbent supplier’s investment, to the benefit of consumers.

However, the temporary monopoly also postpones investment by new suppliers, and so may reduce aggregate investment in both the short run and the long run. Furthermore, investment by new entrants may spur retaliatory or defensive effort by the incumbent supplier.85 This added impetus for investment by an incumbent supplier is eliminated in the short run by a temporary monopoly policy, and potentially reduced by a temporary oligopoly policy. A temporary oligopoly policy also can limit entry to a few selected competitors. Such selection runs the risk of allowing less efficient suppliers to compete, while excluding the most efficient suppliers from the market. A priori

82 Unassisted entry by the first firm in a market generates positive external effects in that market. The first producer causes consumer surplus to increase and does not affect the profits of other producers.

83 The information required to assess the magnitude of relevant external effects includes the details of consumer demand, the extent and nature of product heterogeneity, and the production technologies of actual and potential producers. See Mankiw and Whinston (1986) for additional analysis of this issue.

84 Recall that British Gas was permitted to retain its monopoly status (albeit in continually shrinking market segments) during the initial stages of liberalization.

85 See Woroch (2000) for evidence of this effect.
restrictions on the number of competitors also can limit product innovation, facilitate collusion among the selected incumbent producers, and limit industry pressure to reduce operating costs and prices.

A temporary duopoly is sometimes suggested as a means to assist at least one additional industry participant on the grounds that no entry would occur if an entrant were not promised the security of a temporary duopoly. This rationale is not entirely compelling. If a second entrant would render the first entrant unprofitable, the second entrant seems unlikely to enter unless it is more efficient than the first entrant. Furthermore, an incumbent producer may benefit greatly from entry assistance of this type, which suggests such assistance may be poorly targeted.

An additional potential drawback to a temporary monopoly is that it may provide ample time for an incumbent supplier to devise ways to limit the likely success of future competitors. In addition, “temporary” monopolies can become permanent or semi-permanent monopolies in settings where the incumbent supplier has substantial political power and where the regulator’s commitment powers are limited.

The welfare losses from temporary monopolies and oligopolies can be substantial. In the United States, for example, service quality increased and prices for wireless telecommunications services declined substantially once the relevant markets were opened to more than two providers. During the mandated duopoly period, however, industry prices remained close to their monopoly levels. Similarly, the prices of wireline telecommunications services did not decline substantially in Australia during its mandated duopoly period. In contrast, prices dropped significantly in countries (like Chile and Guatemala, for example) that did not adopt duopoly policies.

6.2 Exclude Foreign Investors.

Feelings of nationalism and concerns with national security can lead to the imposition of limits on foreign participation in key domestic industries. Although such limits can reduce foreign control of domestic industry, they can impose significant costs on the domestic economy. Limits on foreign ownership can reduce the flow of much-needed capital to the domestic industry. The limits can also serve to exclude the most knowledgeable and experienced operators from the industry, and

---

87 See Parker and Röller (1997).
88 See Spiller and Cardilli (1997), for example. The UK adopted a 7 year duopoly policy in its telecommunications industry. Newbery (1999, table 7.4) reports that the productivity of the incumbent producer in the UK improved significantly only when the duopoly policy was relaxed.
89 These concerns may be particularly strong in countries where colonial rule has only recently ended.
thereby reduce industry performance. In cases where operating licenses are sold by a domestic government, limits on foreign participation can also reduce the license revenue that accrues to the domestic government.

6.3 Specify Market Share Targets.

In an attempt to ensure adequate competitive pressure in a liberalized industry, regulators sometimes specify the market share they would like to see competitors achieve.\textsuperscript{90} Such a policy suffers from at least two major drawbacks. First, market share is not necessarily a good measure of market power. In particular, an incumbent supplier may have substantial ability to raise the prices it charges for its services even when competitors presently serve a significant portion of customers. This ability to raise prices profitably may stem from the incumbent’s superior products or control over key inputs, for example. Consequently, the specification of a target market share for competitors may not ensure that substantial discipline is imposed on the incumbent supplier.\textsuperscript{91}

Second, the specification of market share targets can reduce the intensity of market competition. An incumbent supplier may choose not to compete aggressively against a rival in order to ensure that the rival achieves its market share target. When market share targets are specified, the incumbent may rationally refrain from aggressive competition, recognizing that it may cause the regulator to impose more stringent regulations on the incumbent when competitors fail to achieve their market share targets. The resulting diminution in competitive intensity can result in higher prices and lower service quality for customers, and can enable competitors to survive in the marketplace even though they have higher costs than the incumbent supplier.


Many network industries entail a vertical structure that contains a massive network infrastructure segment (e.g., a basic telecommunications network or a gas or electricity transmission system). The incumbent supplier of retail services often owns or operates the infrastructure segment. Given the prohibitive cost of constructing multiple infrastructures, rival retail operators often are compelled to procure key inputs (e.g., network access) from the incumbent, vertically-integrated supplier. The appropriate design of the terms that govern competitors’ access to the incumbent’s network poses a formidable challenge for regulators. Some undesirable (but not

\textsuperscript{90} Recall that target market shares were specified in the UK natural gas industry.

\textsuperscript{91} It is also possible that an incumbent’s market share might over-state the firm’s market power. For instance, consider Bertrand price competition in a setting where one firm has slightly lower unit cost than its rivals. In equilibrium, this firm will have a 100% market share, even though it has limited ability to raise the market price.
uncommon) elements of network access policy are mentioned here. More desirable elements of network access policy are reviewed in section 7.

One inappropriate element of network access policy is a vague or incomplete statement of the incumbent producer’s obligations to supply access to rivals during the liberalization process. If, for example, an incumbent producer is afforded substantial latitude in setting the terms and conditions of network access, the incumbent should be expected to employ this latitude to disadvantage retail rivals. For example, the incumbent should be expected to set high access prices, limit the quality of the inputs delivered to rivals, and delay the provision of access.92

A second inappropriate element of network access policy is the failure to establish a timely, functional dispute resolution process. When they operate as rivals at the retail stage of production, the incumbent producer and retail competitors naturally will have divergent interests regarding the terms and conditions of access. These divergent interests often will lead to disputes, even when the incumbent’s obligations are reasonably well specified. Prompt resolution of these disputes is necessary to ensure the timely implementation of the liberalization process. By specifying clearly the details of a rapid dispute resolution mechanism, a regulator can help to ensure that an incumbent producer does not disadvantage competitors by delaying unduly their access to key network infrastructure.93

A third inappropriate element of network access policy is the implementation of terms of access that are unduly generous for entrants. To reduce the costs that new suppliers incur when entering a market, regulators may afford new suppliers access to essential infrastructure at prices that are below the incumbent’s costs of supplying the access. Short-term subsidies of this nature can reduce the costs that new suppliers incur as they begin to provide service in a formerly monopolized industry. In doing so, such subsidies may help to attract new competitors to the industry and thereby

92 Recall that competing suppliers of natural gas enjoyed little success until the terms of access to British Gas’ infrastructure were explicitly regulated. Weisman (1995), Economides (1998), Reiffen (1998), Mandy (2000), Beard et al. (2001), and Mandy and Sappington (2003), among others, analyze the incentives of a vertically-integrated supplier to disadvantage retail competitors. Of course, a vertically-integrated incumbent supplier may not wish to disadvantage a retail competitor if the competitor is substantially more efficient than the incumbent and if the incumbent can secure sufficiently high profit margins on the inputs it sells to the retail competitor. See, for example, Sibley and Weisman (1998) and Rey and Tirole (2005).

93 Recall that few interconnection agreements were negotiated in Chile until after strict mandates for timely negotiation and an effective dispute resolution mechanism were implemented. Spiller and Cardilli (1997) explain how final-offer arbitration of the type implemented in Guatemala’s telecommunications industry can facilitate timely negotiation of interconnection agreements.
“jump start” competition.\textsuperscript{94}

While short-term subsidies could, in principle, prove beneficial in this regard, subsidies that are intended to be in effect for only a short period of time often remain in effect far longer. Long-term subsidies of this sort introduce at least two important problems. First, the subsidies may permit inefficient firms to operate profitably in the industry, thereby increasing industry costs and reducing industry welfare. Second, subsidized access to infrastructure can induce competitors to employ inefficient operating technologies. In particular, a competitor may decide to employ the subsidized access to the incumbent’s infrastructure even though the competitor would employ fewer social resources if it built and employed its own infrastructure. Thus, by distorting the technological choices of competitors, subsidized infrastructure access can increase industry operating costs. It can also reduce product and process innovation in the industry by limiting the extent to which competitors construct their own infrastructures.\textsuperscript{95}

6.5 Restrict the Incumbent Asymmetrically.

Stringent, asymmetric regulation of an incumbent supplier that limits the incumbent’s ability to compete against entrants can help to attract entry. However, even when entry is desirable, the costs of asymmetric regulation often outweigh its benefits.

To illustrate this more general principle, consider a policy that embeds cross subsidies in the incumbent supplier’s pricing structure, and prohibits the incumbent from reducing any of its prices in response to competitive pressures (even prices that exceed production costs by a substantial margin). Such a policy has at least four potential drawbacks. First, the cream-skimming induced by cross-subsidies can jeopardize the financial integrity of the incumbent supplier. It can do so by reducing the incumbent’s sales of the profitable services targeted by competitors without any offsetting reduction in the sales of the unprofitable, subsidized services. Second, the cross subsidies can increase industry costs by allowing inefficient suppliers to serve customers. Third, when competitors focus their efforts on selling the most profitable services, customers of unprofitable, subsidized services are denied the benefits of competition.\textsuperscript{96} Fourth, as explained more fully in section 7.7 below, the design of appropriate access charges can be complicated considerably when...

\textsuperscript{94} See Gans and King (2004) for an analysis of this issue. Of course, if access prices are unduly high, competitors may build their own infrastructure even though short-term and long-term industry costs would be minimized if the competitors employed the incumbent’s infrastructure.

\textsuperscript{95} See, for example, Hausman (1997), Hausman and Sidak (1999, 2005), Rosston and Noll (2002), and the discussion in section 7.7 below.
the incumbent supplier’s retail tariffs do not reflect the supplier’s costs.

As one further illustration, consider an asymmetric policy that precludes an incumbent supplier from expanding its product line (e.g., an incumbent supplier of telephone service might be precluded from providing cable television service). It is apparent that such regulations can lead to higher prices, reduced product variety, and lower product quality. When a strong, viable competitor is precluded from a market, the remaining operators may not be compelled to compete as aggressively as they otherwise might. Consequently, prices may rise, superior products may be introduced less frequently, and the quality of existing products may decline. Consumers are harmed by each of these outcomes.

In summary, although restrictions on the scope of an incumbent supplier’s operations do not necessarily reduce welfare (see section 7.6 below), they have substantial potential to do so if they are not applied judiciously. Similarly, many of the other types of liberalization considered in this section can, in principle, increase welfare under some conditions. However, these liberalization policies generally are not the best way to promote vigorous, long-term industry competition. Preferable policies are considered next.

7. Pro-Competitive Liberalization Policies.

The purpose of this section is to describe liberalization policies that generally can help to foster vigorous long-term industry competition while avoiding many of the potential drawbacks discussed in section 6. The liberalization policies discussed in this section focus more on removing entry barriers and unleashing the full force of competition rather than on handicapping or favoring certain competitors. Such policies provide increased potential for ultimately relying on market forces rather than on ongoing detailed regulatory oversight to ensure that consumers are well served.

7.1 Reduce Customer Switching and Search Costs.

Competition can compel providers to deliver high-quality products to consumers at low prices if consumers are able to easily identify and secure service from the firms that offer the best products.

---

96 Recall that uniform retail prices and distance-based access charges to British Gas’ transport pipeline may encourage competitors to serve customers located close to gas landing points.

97 Service quality may also decline when universal service obligations require incumbent suppliers to offer service at uneconomic rates. Reductions in service quality can both reduce the cost of supplying the uneconomic service and reduce customer demand for the service. Both effects increase the profit of an incumbent supplier that is saddled with uneconomic universal service obligations.
at the lowest prices.\textsuperscript{98} Therefore, liberalization policies that help to ensure consumers are well informed and are able to switch their service provider easily can stimulate vibrant, enduring competition that may ultimately substitute for regulatory oversight.\textsuperscript{99} Specific policies that can be helpful in this regard include the following three.

First, consumers might be afforded ready access to information about the services that competitors offer. Relevant information includes both price and (objective, verifiable) quality information. If consumers generally have access to the Internet, relevant information might be made available at a government (or government accredited) web site. The web site address might be printed on the bills that customers receive from the incumbent supplier. Price comparisons will be most transparent when the services the firms supply are fairly homogenous (such as gas and electricity). Price comparisons can be less meaningful when industry suppliers offer heterogeneous services (e.g., certain types of telecommunications services). Even when it is impractical to provide comparative information, though, basic information about how to contact competing suppliers can increase consumer awareness of competitive alternatives, and thereby enhance industry competition.\textsuperscript{100}

Second, steps might be taken to reduce the costs that customers incur when they switch suppliers.\textsuperscript{101} For example, providers of local telephone service might be required to install technologies that allow a customer to retain the same telephone number regardless of the supplier from which the customer secures service. Absent such “number portability”, a consumer might be reluctant to switch suppliers because the switch would require the consumer to inform all friends and associates of her new number, or reprint business cards and stationery on which the number appears, for example.\textsuperscript{102}

Third, policies might be implemented to reduce asymmetries in the costs that consumers incur

\textsuperscript{98} See Wolinsky (1997) for a model in which a regulator often allows firms to compete directly for customers rather than specifying market boundaries because customers can better discern service quality than the regulator.

\textsuperscript{99} Recall that many consumers continue to purchase natural gas from British Gas, even though it tends to charge more for gas than its competitors. This behavior may be explained in part by consumer ignorance regarding the prices charged by competitors.

\textsuperscript{100} Models of unregulated oligopoly (e.g., Varian, 1980) often predict that the average equilibrium price in an industry declines as more consumers become informed about the prices competitors charge. Thus, policies that make some consumers more aware of prevailing prices can benefit all consumers.

\textsuperscript{101} See Farrell and Klemperer (2005) for an overview of the economics of switching costs.

\textsuperscript{102} Of course, the benefits of policies like number portability must be weighed carefully against the associated costs (e.g., the costs of designing and installing the requisite technology).
when they choose different suppliers. For example, carrier pre-selection policies in the telecommunications industry allow customers to designate in advance the long-distance carrier that they would like to complete all of their long-distance telephone calls. When a carrier pre-selection policy is in place, customers do not need to dial additional numbers or otherwise undertake costly or time-consuming activities in order to direct their long-distance telephone business to their preferred carrier. Consequently, new carriers do not need to convince customers to incur substantial costs in order to win their loyalty. The new carriers need only provide higher quality and/or lower prices than incumbent carriers.\footnote{Waterson (2003) discusses the potential merits of requiring customers to choose their supplier annually, along the lines of annual policy renewals that are common in the insurance industry. While such a requirement can reduce the differential costs that customers incur in choosing to receive service from a new entrant rather than continuing to receive service from the incumbent supplier, the policy increases transactions costs for all consumers, not only for those who switch suppliers. The same is true of Chile’s equal access policy, which, recall, requires callers to specify their preferred carrier every time they place a long distance call.}

More generally, liberalization policies that provide objective information to consumers about the options available to them, reduce customer switching costs, and limit any differential costs that consumers must incur to obtain service from their preferred supplier can foster vibrant, long-term industry competition.

7.2 Ensure Adequate Monitoring and Data Reporting.

Accurate information about the activities and capabilities of both incumbent suppliers and new operators is of great value when designing regulatory and liberalization policies. This information is essential in assessing the nature and intensity of industry competition, and thus the extent to which stringent regulatory controls can be relaxed.

Accurate information about the services that competitors supply is important in order to identify the services on which the incumbent supplier might reasonably be afforded more substantial pricing flexibility. Market concentration (as measured by the Herfindahl Index, for example) can sometimes serve as an imperfect indicator of the intensity of market competition. Information about the installed capacity of competitors can be of greater value in assessing both the current and likely future intensity of market competition, as can detailed knowledge of the entry barriers that competitors face.

To ensure the timely availability of the information required to formulate appropriate liberalization policies, it is important to establish data reporting requirements. The requirements should specify clearly what data must be reported, how often it must be reported, and which entities
must report the data. The data reporting requirements should be changed as infrequently as possible, so as to limit the costs imposed on the reporting entities (and to provide a consistent time series of data).104

7.3 Privatize State-Owned Enterprises.

When the incumbent supplier is owned (wholly or primarily) by the government rather than by private investors, privatization of the state-owned enterprise (SOE) can be an important element of a successful liberalization policy. By hardening soft budget constraints and promoting profit maximization, privatization of an SOE (i.e., selling the firm to private investors) can: (i) serve to focus the firm’s efforts on reducing its operating costs and delivering high-quality services to consumers; and (ii) reduce a firm’s incentive and ability to engage in below-cost pricing which can preclude the efficient operation of more efficient competitors.105

In settings where a government’s commitment powers are limited, partial privatization of an SOE may be preferable to full privatization. When the government retains an ownership stake in the firm, the government, like private investors, will suffer financially if it implements policies that reduce the firm’s earnings. Consequently, a promise by the government not to expropriate private investors may be more credible when the firm is partially privatized than when it is fully privatized.106 Widespread domestic distribution of the privately-held shares of the (partially) privatized firm can have a similar effect. In the presence of such widespread domestic ownership of the privatized firm, any policy the government might implement that seriously erodes the earnings of the firm would likely evoke widespread, popular opposition. Fearful of the political ramifications of any such widespread opposition, the government will be reluctant to expropriate the privatized firm.107

7.4 Re-balance Retail Tariffs to Better Reflect Costs.

Vibrant long-term industry competition can also be fostered by re-balancing the prices the incumbent supplier charges for its services. Rate re-balancing occurs when prices are aligned more

104 Regulators may rationally choose to keep some of the reported data confidential. Confidentiality can protect proprietary business plans and limit undesired information sharing among industry participants.

105 Sappington and Sidak (2003) note that an SOE’s reduced focus on profit can increase its incentives to engage in anticompetitive activities. If, for example, an SOE is more concerned with market share than profit, the SOE may be more willing than a profit-maximizing firm to implement below-cost pricing for an extended period of time.

106 See, for example, Perotti and Guney (1993), Perotti (1995), and Bel (2003).

107 See Armstrong et al. (1994), Biais and Perotti (2002), and Bel (2003).
closely with incremental production costs. Ideally, the desired rate re-balancing should be completed at the outset of the liberalization process in order to provide appropriate signals to potential competitors about their likely returns from long-term industry operation. Although prices that diverge from cost can help to achieve desired income distribution, pricing structures that embed cross-subsidies are inappropriate when industry liberalization is under way because they introduce the problems identified in section 6.5.

To avoid these problems, the price charged for each of an incumbent’s services should be set at or above the firm’s incremental cost of providing the service. Such rate re-balancing helps to direct the forces of competition to those services that competitors can supply more efficiently than the incumbent supplier, and allow the incumbent supplier to deliver the services that it can supply more efficiently than its rivals.

In settings where rate re-balancing would require dramatic increases in the prices of certain essential services, strong opposition to the re-balancing may arise. To ameliorate this opposition, rates might be re-balanced gradually over time, even though complete re-balancing at the outset of liberalization is preferable, when it is feasible. Alternatively, or in addition, financial support might be provided directly to the customers that would find the price increases to be most burdensome. Such a policy can replace implicit subsidies to all consumers with explicit subsidies to those with the greatest need for financial assistance. If the country’s tax system is relatively efficient, general tax revenue can be the best source of funds for the explicit, targeted subsidies in the regulated industry. If the subsidies must be funded entirely within the regulated industry, taxes can be imposed on services that are deemed to be less essential (e.g., the services whose prices were set well above the incumbent’s costs prior to liberalization). Any such commodity taxes should apply symmetrically to all suppliers, so as not to distort the competitive process.

7.5 Allow Adequate, But Not Unlimited, Pricing Flexibility.

If it has no freedom to change the prices it charges for its services, the incumbent supplier will have limited ability to respond to the challenges presented by competitors. Consequently,

---

108 In the presence of scale and scope economies, it may not be possible to set tariffs that reflect marginal production costs without inflicting a financial deficit on the incumbent supplier.

109 Recall that Chile implemented a program to subsidize the purchase of basic telephone service by citizens in rural regions and low-income urban areas (Laffont, 2005, p. 216).

110 See Atkinson and Stiglitz (1976) and Laffont and Tirole (2000, chapter 6).

111 Armstrong (2002, section 2.1) discusses the design and implementation of such taxes.
competitors may survive in the market even if they are less efficient than the incumbent supplier.\textsuperscript{112}

Price cap regulation can provide incumbent suppliers some pricing flexibility while limiting undue exercise of market power. Price cap regulation plans typically constrain the rate at which the regulated firm’s prices can rise on average, without specifying the exact price that must be charged for any particular service. The application of distinct price cap constraints to distinct groups of services can afford the incumbent supplier some flexibility to respond to competitive challenges without allowing the firm to abuse its market power. To illustrate, the firm might be permitted to increase by five percent annually the average of the prices it charges for services that are supplied by competitors. The firm might also be allowed to increase by one percent annually the average of the prices it charges for services that generally are not supplied by competitors. Pricing restrictions like these that impose different and separate constraints on different groups of services can provide adequate protection for customers who have no competitive alternatives while affording the incumbent supplier a reasonable opportunity to respond to emerging competitive challenges.\textsuperscript{113}

In addition, incumbents should generally be precluded from pricing an established service below the incremental cost of providing the service (unless they are required to do so as part of a universal service obligation). As explained above, below-cost pricing can preclude the operation of more efficient suppliers, and thereby raise industry costs.\textsuperscript{114} Below-cost pricing also encourages excessive consumption (i.e., consumption beyond the point at which the marginal benefit of consumption is equal to the marginal cost of production).

It is particularly important to monitor and implement procedures to preclude below-cost pricing by a firm that operates under price cap regulation. This is because such a firm can have a

\textsuperscript{112} Recall that some observers believe the success enjoyed by new suppliers of local telecommunications services in Chile may be due in part to the limits placed on CTC’s ability to reduce prices selectively in response to competitors’ prices (Paredes, 2005).

\textsuperscript{113} Armstrong et al. (1994, section 7.5.4) describe how British Telecom used the pricing flexibility it was afforded by price cap regulation to reduce prices substantially on long distance telephone calls (that were also supplied by competitors) while raising prices for local telephone calls (for which there were essentially no other suppliers).

\textsuperscript{114} In settings where an incumbent, vertically-integrated producer (VIP) otherwise has considerable pricing flexibility, the relevant price floor can include an “imputation” requirement in order to prevent a price squeeze. A price squeeze occurs when a VIP sets a price for a key input that exceeds the price of the VIP’s relevant retail service. An imputation requirement obligates the VIP to price its retail service above the sum of the VIP’s (downstream) unit cost of producing the retail service and the unit price the VIP charges its retail competitors for the key input. Thus, the lower bound on the price of the VIP’s retail product is the usual incremental cost floor, with one exception. The VIP’s cost of producing the key input is imputed as the input cost that the VIP imposes on its retail competitors. This price floor is thus the incremental cost the VIP would incur if it faced the same input cost as its rivals. This imputation procedure avoids price squeezes, and thereby helps to ensure that the most efficient providers supply the retail service.
particularly strong incentive to price below cost the services on which it faces particularly intense competition. Under a price cap plan that restricts the rate at which all of the firm’s prices can increase on average, a substantial reduction in the price of one service can authorize a pronounced increase in the price of another service. By permitting increases in net revenue on some services that offset corresponding reductions from below-cost pricing on other services, price cap regulation plans can enhance the incentives of an incumbent regulated firm to set predatory prices that can serve to drive its rivals from the market. Therefore, safeguards against below-cost pricing are an important component of price cap regulation plans.

7.6 Prevent Disadvantaging of Downstream Competitors.

Vertically-integrated producers (VIPs) that both sell essential inputs to retail producers and supply the retail service themselves can disadvantage their rivals through anticompetitive actions other than predatory pricing. For example, consider a setting where a VIP operates under price cap regulation. Suppose the price cap regulation plan imposes a single over-arching restriction on the prices of all of the VIP’s services, both retail and wholesale services. Under such a restriction, price reductions on retail services effectively authorize price increases on wholesale services. Both price changes are disadvantageous for retail competitors who must purchase the VIP’s wholesale services. To limit such disadvantaging of rivals, it generally is advisable to place separate restrictions on the rate at which a VIP’s wholesale service prices can rise and on the rate at which its retail service prices can rise.116

VIPs can disadvantage retail competitors through means other than strategic pricing. For example, as indicated in section 6.4, a VIP could intentionally reduce the quality of the inputs it delivers to its downstream competitors, and thereby limit the competitors’ ability to deliver high quality services to their customers.117 To limit undesirable strategic disadvantaging of rivals, careful monitoring of the quality of the inputs that a VIP delivers to its downstream rivals can be advisable, as can an explicit schedule of penalties for unduly low service quality and, perhaps, rewards for exceptionally high levels of service quality.118 Structural separation of the VIP’s wholesale and retail operations can facilitate accurate monitoring of the relative prices and qualities of the inputs

116 Laffont and Tirole (1996) note that when wholesale and retail services are treated as separate baskets of services in this manner, a VIP’s ability and incentive to implement Ramsey prices may be reduced.
117 Reiffen et al. (2000) and Zimmerman (2003) provide some evidence of this effect.
118 Wood and Sappington (2004) discuss the design of such reward structures.
that the VIP supplies to its own retail affiliates and to competing retailers. In assessing the merits of structural separation, this potential benefit should be weighed against any associated costs, including foregone economies of scope.

A more radical alternative to structural separation is ownership separation. Under structural separation, wholesale and retail operations are physically separated, but common ownership and control of both operations is permitted. In contrast, ownership separation precludes common ownership and control of wholesale and retail operations. While structural separation can help to limit the ability of a VIP to disadvantage its retail competitors, ownership separation can serve to reduce the incentive a producer of key wholesale services might have to disadvantage the (retail) firms that purchase the wholesale services. When it does not provide retail services, the wholesale provider does not secure any direct gains in the retail market from disadvantaging retail producers. In fact, a reduction in the quality of the inputs it sells generally will reduce the demand for these inputs, and thereby reduce the revenues of the wholesale producer. Consequently, by eliminating the key source of potential gains from disadvantaging downstream operators, ownership separation can substantially reduce, if not eliminate, incentives for such disadvantaging.

Of course, the costs of ownership separation must be weighed carefully against its potential benefits. Ownership separation, like structural separation, may sacrifice substantial economies of scope and economies associated with integrated planning of wholesale and retail operations. Ownership separation may also preclude a particularly efficient competitor (the producer of wholesale services) from participating in the retail market, and thereby raise industry costs and retail prices. Ownership separation also can entail substantial divestiture costs in cases where a VIP is already providing both wholesale and retail services when ownership separation is considered. The merits of ownership separation also can depend on the prevailing regulatory institutions. A regulator with ample resources, considerable experience, and substantial authority to collect relevant data may be able to prevent a VIP from disadvantaging its rivals. In contrast, a regulator with limited resources, limited experience, and limited information might be unable to do so.

---

119 Hausman and Sidak (2005) review the form of structural separation under consideration by Ofcom, the UK communications regulator. The separation includes the creation of a separate division of British Telecom to supply wholesale services on the same terms and conditions to all retail competitors and an “equality of access” board to oversee the operations of this division.

120 Vertical divestiture of the incumbent also can simplify the design of access charges by eliminating the effects of access charges on the incentives of a VIP to disadvantage retail competitors. Beard et al. (2001), among others, consider how access charges affect the incentives of a VIP to disadvantage rivals.

121 See Vickers (1995), Hinton et al. (1998), Lee and Hamilton (1999), Crandall and Sidak (2002), Buehler et al. (2005), and Crew et al. (2005), for example, for analyses of these and related considerations.
rendering vertical separation the best (albeit a costly) way to promote effective competition.

7.7 Establish Appropriate Access Prices.

The prices that retail competitors must pay to access the infrastructure of an incumbent VIP can have a substantial impact on welfare. A comprehensive assessment of the effects of access prices on welfare is complex. However, some of the key effects can be illustrated in the following simple setting.\(^\text{122}\)

Suppose an incumbent VIP incurs marginal cost \(c_1\) in supplying a single retail service and marginal cost \(c_2\) in supplying a single network input to a single retail competitor. This competitor, the entrant, incurs constant unit cost \(c\) in converting one unit of the access service into one unit of its own retail service. (The production technology is fixed proportions, so exactly one unit of the access service is required to produce each unit of the retail service.) The retail services of the incumbent and the entrant are homogeneous, and consumers purchase the product from the firm that sets the lowest retail price.

Suppose the regulator sets retail price \(p\) and access price \(a\) for the incumbent’s services in this setting. Then the entrant will find it profitable to operate if the highest price \((a + p)\) it can charge for its retail service exceeds its unit cost \((a + c)\) of producing the service. Industry production costs are minimized when the least-cost supplier serves all retail customers. The entrant is the least-cost supplier in this setting when \((c_2 + c)\) is less than \(c_1\). Therefore, for fixed regulated retail price \(p\), the following access price ensures the entrant will operate when and only when it is least-cost industry supplier:\(^\text{123}\)

\[
a = c_2 + [p - c_1].
\]

The access price in equation (10) reflects the efficient component pricing rule (ECPR), which states that the access charge should be set equal to the sum of the incumbent VIP’s cost of supplying access \((c_2)\) and the opportunity cost (or lost profit, \(p - c_1\)) the VIP incurs when it loses a unit of business to the entrant.\(^\text{124}\) The ECPR implies that when the VIP’s regulated retail price exceeds its cost of supplying the retail service (so \(p > c_1\)), the access charge should exceed the

---

\(^{122}\) See Armstrong (2002) and Vogelsang (2003) for more detailed surveys of this topic.

\(^{123}\) Notice that this discussion takes the incumbent’s retail tariff as given. Ideally, wholesale and retail prices should be set simultaneously to achieve social goals, following standard Ramsey principles (Laffont and Tirole, 1994).

\(^{124}\) See Willig (1979) and Baumol (1983) for early discussions of this rule.
VIP’s cost of providing access. This increase in the access charge serves to limit the operation of inefficient entrants. To see why, notice that if a VIP is obligated to both supply network access at cost and set the price of its retail service above cost \((p > c_1)\), an inefficient entrant (i.e., one for which \(c_2 + c > c_1\)) might find it profitable to enter the market, raising industry production costs. This inefficient entry can be precluded by raising the access charge above cost by the same amount the retail tariff is held above cost \((p - c_1)\).\(^{125}\) Similarly, if the VIP is required to offer a retail service at a price below cost \((p < c_1)\), network access can be subsidized accordingly to encourage efficient entry into the subsidized retail market.\(^{126}\)

Intuitively, one might view the VIP as paying a unit tax of \(p - c_1\) when it is required to supply the retail service to consumers at unit price \(p\). To ensure profitable entry only by competitors that are more efficient than the VIP, entrants must pay the same tax the VIP pays. The ECPR implements this tax by raising the access charge above the VIP’s cost of supplying access by \(p - c_1\).

This simple analysis requires modification in richer settings. Suppose, for example, the retail services of the entrant and the VIP are not homogeneous. In this case, one unit of supply by the entrant does not necessarily reduce the VIP’s retail output by exactly one unit, and so the ECPR must be adjusted accordingly.\(^{127}\) Alternatively, suppose the VIP’s retail price is not immutable (perhaps because the VIP operates under a price cap regulation plan that affords the firm some discretion in setting retail prices). In this case, the established access charge will affect the VIP’s choice of retail tariff. The ECPR requires additional modification to reflect this interaction.\(^{128}\)

The design of access prices becomes more complicated if an entrant can supply the input itself, by investing in its own infrastructure, for instance. In such a setting, care must be taken to ensure that access charges provide the entrant with appropriate (“make-or-buy”) incentives to supply inputs itself rather than purchase them from the incumbent VIP. Two distinct regulatory instruments would be ideal in such a setting. To deliver appropriate make-or-buy incentives to entrants, the VIP’s

\(^{125}\)This increase in the access charge above the cost of supplying access can help to mitigate the cream-skimming problem discussed in section 7.4 above.

\(^{126}\)When access charges reflect the ECPR, a VIP is compensated for the profit reduction it incurs when it loses a retail customer to a rival supplier. Consequently, access prices that reflect the ECPR can limit a VIP’s incentive to disadvantage retail rivals.

\(^{127}\)Armstrong et al. (1996) provide the details of this and other extensions of the ECPR, along with a more general characterization of optimal access prices.

\(^{128}\)See Armstrong (2002, section 2.6) for further discussion of this issue.
access charge should be set equal to the VIP’s cost \((c_2)\) of providing the input. This policy ensures an efficient pattern of production given that entry takes place. Such cost-based access pricing may not ensure industry cost minimization, though, because inefficient retail competitors may find entry profitable or efficient competitors may find entry unprofitable when retail prices diverge from the VIP’s production costs \((p \neq c_1)\). To ensure efficient entry decisions, cost-based access charges should be accompanied by a tax on the outputs of entrants that reflects the deviation of the VIP’s retail price from its cost \((p - c_1)\).\(^{129}\) When the regulator is unable to implement output taxes, the access charge is forced to perform the dual task of providing efficient make-or-buy decisions and efficient entry decisions.\(^{130}\) Typically, a single instrument cannot achieve two goals, making some compromise inevitable.\(^{131}\)

In summary, the appropriate design of access charges can require considerable information. Furthermore, the delivery of appropriate incentives for efficient make-or-buy and industry participation decisions can require an extensive set of regulatory instruments (e.g., access prices and output taxes). In settings where a regulator has limited information and limited powers, the design of access charges can present extremely challenging problems. However, these problems can be mitigated if the incumbent VIP’s retail tariffs reflect its production costs, as proposed in section 7.4 above. When retail rates reflect costs (so \(p = c_1\)) in the simple setting considered here, the ECPR in equation (10) collapses to a particularly simple rule: the access charge should be set equal to the VIP’s cost of providing access. As noted above, this rule also provides appropriate make-or-buy incentives. Notice that this policy can require little knowledge of consumer demand, and does not require the regulator to be able to control the activities of entrants. In sum, an effective rebalancing of the incumbent VIP’s retail tariff greatly simplifies the regulator’s task of setting appropriate access charges, and allows access charges to focus on the single task of ensuring appropriate make-or-buy decisions.

Access prices also can affect incentives for network innovation and cost reduction. For example, access prices that provide profit margins on access services that vary inversely with the incumbent VIP’s realized access costs can create incentives for the VIP to reduce its costs of

\(^{129}\)See Armstrong (2002, section 2.4) for further discussion of this point.

\(^{130}\)Sappington (2005b) notes that if retail prices are not regulated, efficient make-or-buy decisions may arise even when access charges do not reflect the incumbent’s production costs.

\(^{131}\)Hausman (1997) and Hausman and Sidak (1999, 2005), among others, note that competitors will rely unduly on an incumbent supplier’s infrastructure if access prices are set below the full cost of supplying access. This full cost includes a capital cost that reflects prevailing technological and market risks.
supplying access. Clearly, the appropriate design of access charges becomes more complex when access charges serve to motivate cost reduction in addition to securing industry cost minimization and inducing efficient make-or-buy decisions.

Conceivably, access prices might be set at relatively high levels to systematically encourage entrants to construct their own infrastructure. One reason for doing so (despite the duplicative infrastructure costs such a policy can promote) would be to promote competition among multiple producers, in hopes of fostering industry innovation and eliminating the need for long-term regulation of a monopoly supplier of key inputs. However, competition among facilities-based networks will not necessarily eliminate the need for long-term regulation. Even if all operators in the telecommunications industry employ their own infrastructure, each operator must still interconnect with other operators in order to complete calls intended for the customers of other operators. It is conceivable that unregulated negotiations among similarly situated network operators could produce inter-carrier compensation arrangements that are broadly consistent with social goals. However, this need not be the case. Firms might negotiate compensation arrangements that inhibit competition or limit new entry, for example.

7.8 Limit Cost Shifting.

In settings where regulatory controls are relaxed on some, but not all, of the incumbent supplier’s operations, it can be important to limit the supplier’s ability to engage in cost shifting. Cost shifting occurs when costs that actually are incurred in the production of one set of services (e.g., competitive services) are recorded as costs incurred in the production of a different set of services (e.g., monopoly services). An incumbent supplier may benefit from cost shifting when, for example, its competitive operations are not regulated but the prices it can charge for its monopoly services are linked directly to its measured operating costs. In this situation, the incumbent supplier can secure higher regulated prices without affecting its (pre-tax) earnings on competitive services by shifting costs from its unregulated to its regulated operations.

132 Access prices in the U.S. telecommunications industry reflect estimates of an efficient incumbent supplier would incur in supplying access, which can differ from the actual costs of the incumbent supplier. (See Rosston and Noll (2002), for example.) This pricing policy is intended to induce incumbent suppliers to operate efficiently.

133 See Woroch (2002). Boles de Boer et al. (2000) note that New Zealand encouraged facilities-based competition among Internet service providers while Australia relied more on regulation to protect consumers. The authors report lower prices in New Zealand than in Australia.

The incumbent supplier’s ability to undertake such cost shifting can be limited by requiring separate books of account for its regulated and unregulated operations, for example. The firm’s incentive to engage in cost shifting can be limited by reducing the extent to which regulated prices are linked to measured operating costs. In particular, price cap regulation can reduce incentives for cost shifting relative to rate of return regulation and other forms of cost-plus regulation.135

7.9 Ensure the Integrated Operation of All Elements of Industry Policy.

It is important to review industry policy in its entirety whenever a major component of industry policy (e.g., liberalization policy) is altered substantially. The review should be designed to ensure the ongoing, effective, integrated operation of all components of industry policy.

The experience in California’s electricity industry illustrates the importance of such a comprehensive review. The deregulation of wholesale electricity prices can benefit consumers when a competitive wholesale supply of electricity is available. Encouraging major buyers and suppliers of wholesale electricity to participate in a common exchange system also can provide widespread benefits. Furthermore, a capped retail price for electricity can be a component of a sensible regulatory policy. However, even though each of these policies can have individual merit, the combination of these policies can lead to highly undesirable outcomes. As in California, volatile wholesale prices coupled with capped retail prices can jeopardize the financial integrity of retail suppliers of electricity. The problem can become particularly acute when the volatility of wholesale prices is increased by limiting the ability of retail suppliers to negotiate long term contracts for electricity.

Several of the policies cited above can be viewed as corollaries of the general principle that effective, integrated operation of all components of industry policy merit careful examination whenever liberalization is considered. For example, the need to rebalance retail prices and to afford incumbent suppliers appropriate pricing flexibility reflect prudent changes in certain elements of industry policy as other elements change.

7.10 Increase Antitrust Scrutiny and Enforcement.

A central long-term goal of liberalization is to replace regulatory oversight and control with the disciplining forces of competition. Liberalization might be viewed as a process by which competitive forces are fostered and strengthened to the point where they, alone, can impose effective discipline on some or all of the incumbent supplier’s operations. The more pronounced the

135 Braeutigam and Panzar (1989), Brennan (1990), and Weisman (1993), among others, provide formal analyses of this issue.
disciplinary role played by competition, the more important it is to protect competition and ensure it is not subverted by powerful industry participants. This is generally the role of antitrust policy.

Antitrust policy and regulatory policy differ in at least three fundamental respects. First, antitrust policy typically sets guidelines that describe in broad terms acceptable behavior and outcomes. In contrast, regulatory policy often specifies detailed rules (often that apply to particular firms) that define fairly precisely the limits on acceptable behavior and outcomes. Second, having specified \textit{ex ante} guidelines, antitrust policy typically entails \textit{ex post} investigations of possible violations of the specified guidelines. Regulation, in contrast, often couples \textit{ex ante} rules with ongoing industry oversight, rule refinement, and rule enforcement. Third, antitrust policy typically relies on edicts to discontinue anticompetitive behavior and associated fines (often in the form of damage payments to injured parties), while regulation often proscribes specific types of conduct (e.g., price discrimination or expansion into particular markets), and establishes detailed performance requirements and associated reward and penalty structures.\footnote{See, for example, Borreau and Dogan (2001), Sidak (2003), Cave (2004), and Gerardin and Sidak (2005). Carlton and Picker (2005) note that regulation can allow interested parties greater ongoing intervention in the control process than antitrust.}

Antitrust and regulatory policy can play vital and complementary roles in the liberalization process. Indeed, it may be important to increase both regulatory oversight and antitrust enforcement as competition develops in the liberalized industry. Thus, the road to deregulation of an industry may not be a straight one. Increased antitrust enforcement and regulatory oversight may both be necessary temporarily to ensure that competition has the opportunity to develop to the point where it can eventually replace regulation as the key source of discipline on the incumbent firm.\footnote{Ultimately, explicit regulatory control may be replaced in some settings simply by the specter of regulatory intervention. See Glazer and McMillan (1992), for example.}

In the 1990s, New Zealand decided to rely solely on general competition laws enforced by the courts and by a non-specialized competition authority to govern activities in the telecommunications and electricity sectors. This novel approach was not an immediate failure. However, problems emerged over time, leading the government to realize the need for more orthodox regulatory control. The problems included a heavy case load for the courts and substantial difficulty in proving an incumbent supplier had abused its dominant position. Moreover, legal proceedings proved to move slowly at times when quick decisions were required to keep pace with rapid technological advance. The New Zealand experience suggests that exclusive reliance on competition law may not be sufficient in network industries with emerging competition, even in
8. Conclusions.

The foregoing discussion has produced two broad conclusions. First, even the apparently simple choice between regulated monopoly and unregulated competition is not always straightforward in practice. The appropriate choice between these regimes can vary with the relevant technological and demand conditions, with the regulator’s skills and resources, with the efficiency of tax systems and capital markets, and with the strength of other prevailing institutions. Second, in settings where competition is preferable to regulated monopoly, the road from monopoly to competition can be a particularly long and winding road. There is no single ideal path from monopoly to competition. The most appropriate liberalization policy depends upon a wide variety of factors, including those relevant to the choice between monopoly and regulation.

The foregoing discussion also has emphasized that, although there is no single liberalization policy that is ideal in all settings, some policies typically are superior to others. Liberalization policies that primarily aid some competitors and handicap others on an ongoing basis can hinder the development of vigorous long-term competition. Therefore, policies such as establishing temporary monopolies or oligopolies, excluding foreign investors, specifying market share targets for industry suppliers, providing entrants with long-term subsidized access to the incumbent’s infrastructure, restricting unduly the incumbent supplier’s pricing flexibility, and imposing unfunded carrier-of-last-resort obligations exclusively on incumbent suppliers generally are not recommended.

In contrast, liberalization policies that remove barriers to entry and empower consumers to discipline industry suppliers typically are better methods for fostering vigorous long-term industry competition. In particular, policies like reducing customer switching costs, re-balancing the incumbent supplier’s prices to better reflect its operating costs, privatizing state-owned enterprises, prohibiting below-cost pricing, and establishing appropriate (access) prices for the use of critical infrastructure generally are recommended.

In addition, careful monitoring of industry operations can be of critical importance during the liberalization process. Accurate, timely data about the nature and intensity of industry competition will allow regulatory policy to adjust quickly to changes in industry conditions. Consequently, although liberalization should ultimately lead to reduced regulatory oversight and control, more pronounced regulatory and antitrust oversight may be required on an interim basis to ensure that

---

regulatory policy is tailored appropriately to the evolving level of competition and that competition is protected.

The road from monopoly to competition is seldom straight and smooth. Detours (e.g., increased regulatory and antitrust scrutiny) may be necessary to ensure safe passage to the intended destination. Bumps in the road (e.g., widespread opposition to rate re-balancing or privatization of state-owned enterprises) should be anticipated in order to facilitate their navigation in as smooth a manner as possible. Thus, the liberalization process can constitute a challenging journey, and one that entails considerable uncertainty. Detailed road maps can be of enormous value on such journeys. Unfortunately, though, because every setting in which liberalization might be contemplated is different, road maps that provide sufficient granularity and specificity in every relevant setting are not available. Therefore, one generally must embark on the road to competition armed only with the coarse route markers on which the foregoing discussion has focused.

As emphasized in section 4, the most appropriate liberalization policy can vary considerably according to the institutional setting in which it is being implemented. Therefore, an important role for future research is to develop detailed maps of the best route to competition, i.e., to specify the precise details of liberalization policies that will work well in specific institutional settings. Two approaches to this task seem particularly fruitful. First, the broad principles reviewed in this paper can be translated into precise policy recommendations. (For example, precisely how rapidly should rates be re-balanced, given the prevailing prices, customer wealth, and the operating technologies of the incumbent suppliers and entrants? Alternatively, at precisely what level should access prices be set given the prevailing industry costs structure and retail prices?) Second, case studies and comprehensive empirical work can be undertaken to assess the success and failure of particular liberalization policies in different institutional settings. Together, these two approaches can expand our knowledge of the liberalization policies that will best harness the powers of competition to provide critical industry discipline, and perhaps ultimately replace the costly, imperfect discipline that regulatory oversight and control provide.

References


Hausman, Jerry. 1997. “Valuing the Effect of Regulation on New Services in


