Chapter 27

RECENT DEVELOPMENTS IN THE THEORY OF REGULATION

MARK ARMSTRONG

Department of Economics, University College London

DAVID E.M. SAPPINGTON

Department of Economics, University of Florida

Contents

Handbook of Industrial Organization, Volume 3
Edited by M. Armstrong and R. Porter

DOI: 10.1016/S1573-448X(06)03027-5

Abstract

This chapter reviews recent theoretical work on the design of regulatory policy, focusing on the complications that arise when regulated suppliers have better information about the regulated industry than do regulators. The discussion begins by characterizing the optimal regulation of a monopoly supplier that is better informed than the regulator about its production cost and/or consumer demand for its product. Both adverse selection ("hidden information") and moral hazard ("hidden action") complications are considered, as are the additional concerns that arise when the regulator's intertemporal commitment powers are limited. The chapter then analyzes the design of practical policies, such as price cap regulation, that are often observed in practice. The design of regulatory policy in the presence of limited competitive forces also is reviewed. Yardstick regulation, procedures for awarding monopoly franchises, and optimal industry structuring are analyzed. The chapter also analyzes the optimal pricing of access to bottleneck production facilities in vertically-related industries, stressing the complications that arise when the owner of the bottleneck facility also operates as a retail producer.

Keywords

Regulation, Monopoly, Asymmetric information, Liberalization

JEL classification: D42, D60, D82, L12, L13, L43, L51

1. Introduction

Several chapters in this volume analyze unfettered competition between firms. Such analyses are instrumental in understanding the operation of many important industries. However, activities in some industries are determined in large part by direct government regulation of producers. This is often the case, for example, in portions of the electricity, gas, sanitation, telecommunications, transport, and water industries. This chapter reviews recent analyses of the design of regulatory policy in industries where unfettered competition is deemed inappropriate, often because technological considerations render supply by one or few firms optimal.

The discussion in this chapter focuses on the complications that arise because regulators have limited knowledge of the industry that they regulate. In practice, a regulator seldom has perfect information about consumer demand in the industry or about the technological capabilities of regulated producers. In particular, the regulator typically has less information about such key industry data than does the regulated firm(s). Thus, a critical issue is how, if at all, the regulator can best induce the regulated firm to employ its privileged information to further the broad interests of society, rather than to pursue its own interests.

As its title suggests, this chapter will focus on recent theoretical contributions to the regulation literature.[1] Space constraints preclude detailed discussions of the institutional features of individual regulated industries. Instead, the focus is on basic principles that apply in most or all regulated industries.[2] The chapter proceeds as follows. Section 2 considers the optimal regulation of a monopoly producer that has privileged information about key aspects of its environment. The optimal regulatory policy is shown to vary with the nature of the firm's private information and with the intertemporal commitment powers of the regulator, among other factors. The normative analysis in Section 2 presumes that, even though the regulator's information is not perfect, he is well informed about the structure of the regulatory environment and about the precise manner in which his knowledge of the environment is limited.[3]

Section 3 provides a complementary positive analysis of regulatory policies in a monopoly setting where the regulator's information, as well as his range of instruments, may be much more limited. The focus of Section 3 is on regulatory policies that perform "well" under certain relevant circumstances, as opposed to policies that are optimal in the specified setting. Section 3 also considers key elements of regulatory policies that have gained popularity in recent years, including price cap regulation.

[1] The reader is referred to Baron (1989) and Braeutigam (1989), for example, for excellent reviews of earlier theoretical contributions to the regulation literature. Although every effort has been made to review the major analyses of the topics covered in this chapter, every important contribution to the literature may not be cited. We offer our apologies in advance to the authors of any uncited contribution, appealing to limited information as our only excuse.

[2] We also do not attempt a review of studies that employ experiments to evaluate regulatory policies. For a recent overview of some of these studies, see Eckel and Lutz (2003).

[3] Throughout this chapter, we will refer to the regulator as "he" for expositional simplicity.

Section 4 analyzes the design of regulatory policy in settings with multiple firms. This section considers the optimal design of franchise bidding and yardstick competition. It also analyzes the relative merits of choosing a single firm to supply multiple products versus assigning the production of different products to different firms. Section 4 also explains how the presence of unregulated rivals can complement, or complicate, regulatory policy.

Section 5 considers the related question of when a regulated supplier of a monopoly input should be permitted to compete in downstream markets. Section 5 also explores the optimal structuring of the prices that a network operator charges for access to its network. The design of access prices presently is an issue of great importance in many industries where regulated suppliers of essential inputs are facing increasing competition in the delivery of retail services. In contrast to most of the other analyses in this chapter, the analysis of access prices in Section 5 focuses on a setting where the regulator has complete information about the regulatory environment. This focus is motivated by the fact that the optimal design of access prices involves substantial subtleties even in the absence of asymmetric information.

The discussion concludes in Section 6, which reviews some of the central themes of this chapter, and suggests directions for future research.

2. Optimal monopoly regulation

2.1. Aims and instruments

The optimal regulation of a monopoly supplier is influenced by many factors, including:

1. the regulator's objective (when he is benevolent);
2. the cost of raising revenue from taxpayers;
3. the range of policy instruments available to the regulator, including his ability to tax the regulated firm or employ public funds to compensate the firm directly;
4. the regulator's bargaining power in his interaction with the firm;
5. the information available to the regulator and the firm;
6. whether the regulator is benevolent or self-interested; and
7. the regulator's ability to commit to long-term policies.

The objective of a benevolent regulator is modeled by assuming the regulator seeks to maximize a weighted average of consumer (or taxpayer) surplus, S, and the rent (or net profit), R, secured by the regulated firm. Formally, the regulator is assumed to maximize $S + \alpha R$, where $\alpha \in [0, 1]$ is the value the regulator assigns to each dollar of rent. The regulator's preference for consumer surplus over rent (indicated by $\alpha < 1$) reflects a greater concern with the welfare of consumers than the welfare of shareholders. This might be due to differences in their average income, or because the regulator cares about the welfare of local constituents and many shareholders reside in other jurisdictions.[4]

The second factor – the cost of raising funds from taxpayers – is captured most simply by introducing the parameter $\Lambda \geqslant 0$. In this formulation, taxpayer welfare is presumed to decline by $1 + \Lambda$ dollars for each dollar of tax revenue the government collects. The parameter Λ, often called the *social cost of public funds*, is strictly positive when taxes distort productive activity (reducing efficient effort or inducing wasteful effort to avoid taxes, for example), and thereby create deadweight losses. The parameter Λ is typically viewed as exogenous in the regulated industry.[5]

The literature generally adopts one of two approaches. The first approach, which follows Baron and Myerson (1982), abstracts from any social cost of public funds (so $\Lambda = 0$) but presumes the regulator strictly prefers consumer surplus to rent (so $\alpha < 1$). The second approach, which follows Laffont and Tirole (1986), assumes strictly positive social costs of public funds (so $\Lambda > 0$) but abstracts from any distributional preferences (so $\alpha = 1$). The two approaches provide similar qualitative conclusions, as does a combination of the two approaches (in which $\Lambda > 0$ and $\alpha < 1$). Therefore, because the combination introduces additional notation that can make the analysis less transparent, the combination is not pursued here.[6]

The central difference between the two basic approaches concerns the regulated prices that are optimal when the regulator and firm are both perfectly informed about the industry demand and cost conditions. In this benchmark setting, the regulator who faces no social cost of funds will compensate the regulated firm directly for its fixed costs of production and set marginal-cost prices. In contrast, the regulator who finds it costly to compensate the firm directly (since $\Lambda > 0$) will establish Ramsey prices, which exceed marginal cost and thereby secure revenue to contribute to public funds. Because the marginal cost benchmark generally facilitates a more transparent analysis, the ensuing analysis will focus on the approach in which the regulator has a strict preference for consumer surplus over rent but faces no social cost of public funds.

The third factor – which includes the regulator's ability to compensate the firm directly – is a key determinant of optimal regulatory policy. The discussion in Section 2 will follow the strand of the literature that presumes the regulator can make direct payments to the regulated firm. In contrast, the discussion of practical policies in Section 3 generally will follow the literature that assumes such direct payments are not feasible (because the regulator has no access to public funds, for example).

The fourth factor – the regulator's bargaining power – is typically treated in a simple manner: the regulator is assumed to possess all of the bargaining power in his interaction

[4] Baron (1988) presents a positive model of regulation in which the regulator's welfare function is determined endogenously by a voting process.

[5] In general, the value of Λ is affected by a country's institutions and macroeconomic characteristics, and so can reasonably be viewed as exogenous to any particular regulatory sector. Laffont (2005, pp. 1–2) suggests that Λ may be approximately 0.3 in developed countries, and well above 1 in less developed countries.

[6] As explained further below, the case where $\alpha = 1$ and $\Lambda = 0$ is straightforward to analyze.

with the regulated firm. This assumption is modeled formally by endowing the regulator with the ability to offer a regulatory policy that the firm can either accept or reject. If the firm rejects the proposed policy, the interaction between the regulator and the firm ends. This formulation generally is adopted for technical convenience rather than for realism.[7]

The fifth factor – the information available to the regulator – is the focus of Section 2. Regulated firms typically have better information about their operating environment than do regulators. Because of its superior resources, its ongoing management of production, and its frequent direct contact with customers, a regulated firm will often be better informed than the regulator about both its technology and consumer demand. Consequently, it is important to analyze the optimal design of regulatory policy in settings that admit such adverse selection (or "hidden information") problems.

Two distinct adverse selection problems are considered in Section 2.3. In the first setting, the firm is better informed than the regulator about its operating cost. In the second setting, the firm has privileged information about consumer demand in the industry. A comparison of these settings reveals that the properties of optimal regulatory policies can vary substantially with the nature of the information asymmetry between regulator and firm. Section 2.3 concludes by presenting a unified framework for analyzing these various settings.

Section 2.4 provides some extensions of this basic model. Specifically, the analysis is extended to allow the regulator to acquire better information about the regulated industry, to allow the firm's private information to be multi-dimensional, and to allow for the possibility that the regulator is susceptible to capture by the industry (the sixth factor listed above). Section 2.5 reviews how optimal regulatory policy changes when the interaction between the regulator and firm is repeated over time. Optimal regulatory policy is shown to vary systematically according to the regulator's ability to make credible commitments to future policy (the seventh factor cited above).

Regulated firms also typically know more about their actions (e.g., how diligently managers labor to reduce operating costs) than do regulators. Consequently, it is important to analyze the optimal design of regulatory policy in settings that admit such moral hazard (or "hidden action") problems. Section 2.6 analyzes a regulatory moral hazard problem in which the firm's cost structure is endogenous.

2.2. Regulation with complete information

Before analyzing optimal regulatory policy when the firm has privileged knowledge of its environment, consider the full-information benchmark in which the regulator is omniscient. Suppose a regulated monopolist supplies n products. Let p_i denote the price of

[7] Bargaining between parties with private information is complicated by the fact that the parties may attempt to signal their private information through the contracts they offer. Inderst (2002) proposes the following alternative to the standard approach in the literature. The regulator first makes an offer to the (better informed) monopolist. If the firm rejects this offer, the firm can, with some exogenous probability, respond with a final take-it-or-leave-it offer to the regulator.

product i, and let $\mathbf{p} = (p_1, \ldots, p_n)$ denote the corresponding vector of prices. Further, let $v(\mathbf{p})$ denote aggregate consumer surplus and $\pi(\mathbf{p})$ denote the monopolist's profit with price vector $\mathbf{p}$. The important difference between the analysis here and in the remainder of Section 2 is that the regulator knows the functions $v(\cdot)$ and $\pi(\cdot)$ perfectly here.

2.2.1. *Setting where transfers are feasible*

Consider first the setting where the regulator is able to make transfer payments to the regulated firm and receive transfers from the firm. Suppose the social cost of public funds is Λ. To limit the deadweight loss from taxation in this setting, the regulator will extract all the firm's rent and pass it onto taxpayers in the form of a reduced tax burden. (The regulator's relative preference for profit and consumer surplus, i.e., the parameter α, plays no role in this calculation.) Since a \$1 reduction in the tax burden makes taxpayers better off by \$$(1 + \Lambda)$, total welfare with the price vector $\mathbf{p}$ is

$$v(\mathbf{p}) + (1 + \Lambda)\pi(\mathbf{p}). \tag{1}$$

A regulator should choose prices to maximize expression (1) in the present setting. In the special case where $\Lambda = 0$ (as presumed in the remainder of Section 2), prices will be chosen to maximize total surplus $v + \pi$. Optimal prices in this setting are marginal-cost prices. This ideal outcome for the regulator will be called the *full-information outcome* in the ensuing analysis. When $\Lambda > 0$, prices optimally exceed marginal costs (at least on average). For instance, in the single-product case, the price p that maximizes expression (1) satisfies the Lerner formula

$$\frac{p - c}{p} = \left[\frac{\Lambda}{1 + \Lambda}\right]\frac{1}{\eta(p)}, \tag{2}$$

where c is the firm's marginal cost and $\eta = -pq'(p)/q(p)$ is the elasticity of demand for the firm's product (and where a prime denotes a derivative).

2.2.2. *Setting where transfers are infeasible*

Now consider the setting in which the regulator cannot use public funds to finance transfer payments to the firm and cannot directly tax the firm's profit. Because the regulator cannot compensate the firm directly in this setting, the firm must secure revenues from the sale of its products that are at least as great as the production costs it incurs. When the firm operates with increasing returns to scale, marginal-cost prices will generate revenue below cost. Consequently, the requirement that the firm earn non-negative profit will impose a binding constraint on the regulator, and the regulator will choose prices to maximize total surplus $(v(\mathbf{p}) + \pi(\mathbf{p}))$ while ensuring zero profit for the firm ($\pi(\mathbf{p}) = 0$). This is the Ramsey–Boiteux problem.[8] (Again, the regulator's relative preference for profit and consumer surplus plays no meaningful role here.) In the single-product case, the optimal policy is to set a price equal to the firm's average cost of production. If we let λ denote the Lagrange multiplier associated with the break-even constraint ($\pi(\mathbf{p}) = 0$), then under mild regularity conditions, Ramsey–Boiteux prices maximize $v(\mathbf{p}) + (1 + \lambda)\pi(\mathbf{p})$, which has the same form as expression (1). Thus, optimal prices in the two problems – when transfers are possible and there is a social cost of public funds, and when transfers are not possible – take the same form. The only difference is that the "multiplier" Λ is exogenous to the former problem, whereas λ is endogenous in the latter problem and chosen so that the firm earns exactly zero profit.[9]

2.3. Regulation under adverse selection

In this section we analyze simple versions of the central models of optimal regulation with private, but exogenous, information.[10] The models are first discussed under the headings of private information about cost and private information about demand. The ensuing discussion summarizes the basic insights in a unified framework.

2.3.1. Asymmetric cost information

We begin the discussion of optimal regulatory policy under asymmetric information by considering an especially simple setting. In this setting, the regulated monopoly sells one product and customer demand for the product is known precisely to all parties. In particular, the demand curve for the regulated product, $Q(p)$, is common knowledge, where $p \geqslant 0$ is the unit price for the regulated product. The only information asymmetry concerns the firm's production costs, which take the form of a constant marginal cost c together with a fixed cost F. Three variants of this model are discussed in turn. In the first variant, the firm has private information about its marginal cost alone, and this cost is exogenous and is not observed by the regulator. In the second variant, the firm is privately informed about both its fixed and its marginal costs of production. The regulator knows the relationship between the firm's exogenous marginal and fixed costs, but cannot observe either realization. In the third variant, the firm can control its marginal cost and the regulator can observe realized marginal cost, but the regulator is

[8] Ramsey (1927) examines how to maximize consumer surplus while employing proportional taxes to raise a specified amount of tax revenue. Boiteux (1956) analyzes how to maximize consumer surplus while marking prices up above marginal cost to cover a firm's fixed cost.

[9] In the special case where consumer demands are independent (so there are no cross price effects), Ramsey prices follow the "inverse elasticity" rule, where each product's Lerner index $(p_i - c_i)/p_i$ is inversely proportional to that product's elasticity of demand. More generally, Ramsey prices have the property that an amplification of the implicit tax rates $(p_i - c_i)$ leads to an equi-proportionate reduction in the demand for all products [see Mirrlees (1976), Section 2].

[10] For more extensive and general accounts of the theory of incentives, see Laffont and Martimort (2002) and Bolton and Dewatripont (2005), for example.

not fully informed about the fixed cost the firm must incur to realize any specified level of marginal cost.

In all three variants of this model, the regulator sets the unit price p for the regulated product. The regulator also specifies 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's rent, R, is its profit, $\pi = Q(p)(p - c) - F$, plus the transfer T it receives.

Unknown marginal cost[11] For simplicity, suppose the firm produces with constant marginal cost that can take one of two values, $c \in \{c_L, c_H\}$. Let $\Delta^c = c_H - c_L > 0$ denote the cost differential between the high and low marginal cost. The firm knows from the outset of its interaction with the regulator whether its marginal cost is low, c_L, or high, c_H. The regulator does not share this information, and never observes cost directly. He views marginal cost as a random variable that takes on the low value with probability $\phi \in (0, 1)$ and the high value with probability $1 - \phi$. In this initial model, it is common knowledge that the firm must incur fixed cost $F \geqslant 0$ in order to operate.

In this setting, the full-information outcome is not feasible. To see why, suppose the regulator announces that he will implement unit price p_i and transfer payment T_i when the firm claims to have marginal cost c_i, for $i = L, H$.[12] When the firm with cost c_i chooses the (p_i, T_i) option, its rent will be

$$R_i = Q(p_i)(p_i - c_i) - F + T_i. \tag{3}$$

In contrast, if this firm chooses the alternative (p_j, T_j) option, its rent is

$$Q(p_j)(p_j - c_i) - F + T_j = R_j + Q(p_j)(c_j - c_i).$$

It follows that if the low-cost firm is to be induced to choose the (p_L, T_L) option, it must be the case that

$$R_L \geqslant R_H + \Delta^c Q(p_H). \tag{4}$$

Therefore, the full-information outcome is not feasible, since inequality (4) cannot hold when both $R_H = 0$ and $R_L = 0$.[13]

To induce the firm to employ its privileged cost information to implement outcomes that approximate (but do not replicate) the full-information outcome, the regulator pursues the policy described in Proposition 1.[14]

[11] This discussion is based on Baron and Myerson (1982). The qualitative conclusions derived in our simplified setting hold more generally. For instance, Baron and Myerson derive corresponding conclusions in a setting with non-linear costs where the firm's private information is the realization of a continuous random variable.

[12] The revelation principle ensures that the regulator can do no better than to pursue such a policy. See, for example, Myerson (1979) or Harris and Townsend (1981).

[13] This conclusion assumes it is optimal to produce in the high-cost state. This assumption will be maintained throughout the ensuing discussion, unless otherwise noted.

[14] A sketch of the proofs of Propositions 1 through 4 is provided in Section 2.3.3.

PROPOSITION 1. *When the firm is privately informed about its marginal cost of production, the optimal regulatory policy has the following features*:

$$p_L = c_L; \qquad p_H = c_H + \frac{\phi}{1-\phi}(1-\alpha)\Delta^c; \tag{5}$$

$$R_L = \Delta^c Q(p_H); \qquad R_H = 0. \tag{6}$$

As expression (6) reveals, the regulator optimally provides the low-cost firm with the minimum possible rent required to ensure it does not exaggerate its cost. This is the rent the low-cost firm could secure by selecting the (p_H, T_H) option. To reduce this rent, p_H is raised above c_H. The increase in p_H reduces the output of the high-cost firm, and thus the number of units of output on which the low-cost firm can exercise its cost advantage by selecting the (p_H, T_H) option. (This effect is evident in inequality (4) above.) Although the increase in p_H above c_H reduces the rent of the low-cost firm – which serves to increase welfare when c_L is realized – it reduces the total surplus available when the firm's cost is c_H. Therefore, the regulator optimally balances the expected benefits and costs of raising p_H above c_H. As expression (5) indicates, the regulator will set p_H further above c_H the more likely is the firm to have low cost (i.e., the greater is $\phi/(1-\phi)$) and the more pronounced is the regulator's preference for limiting the rent of the low-cost firm (i.e., the smaller is α).

Expression (5) states that the regulator implements marginal-cost pricing for the low-cost firm. Any deviation of price from marginal cost would reduce total surplus without any offsetting benefit. Such a deviation would not reduce the firm's expected rent, since the high-cost firm has no incentive to choose the (p_L, T_L) option. As expression (6) indicates, the firm is effectively paid only c_L per unit for producing the extra output $Q(p_L) - Q(p_H)$, and this rate of compensation is unprofitable for the high-cost firm.

Notice that if the regulator valued consumer surplus and rent equally (so $\alpha = 1$), he would not want to sacrifice any surplus when cost is c_H in order to reduce the low-cost firm's rent. As expression (5) shows, the regulator would implement marginal-cost pricing for both cost realizations. Doing so would require that the low-cost firm receive a rent of at least $\Delta^c Q(c_H)$. But the regulator is not averse to this relatively large rent when he values rent as highly as consumer surplus.

This last conclusion holds more generally as long as the regulator knows how consumers value the firm's output.[15] To see why, write $v(p)$ for consumer surplus when the price is p, and write $\pi(p)$ for the firm's profit function (a function that may be known only by the firm). Suppose the regulator promises the firm a transfer of $T = v(p)$ when it sets the price p. Under this reward structure, the firm chooses its price to maximize $v(p) + \pi(p)$, which is just social welfare when $\alpha = 1$. The result is marginal-cost pricing. In effect, this policy makes the firm the residual claimant for social surplus, and

[15] See Loeb and Magat (1979). Guesnerie and Laffont (1984) also examine the case where the regulator is not averse to the transfers he delivers to the firm.

thereby induces the better-informed party to employ its superior information in the social interest. Such a policy awards the entire social surplus to the firm. However, this asymmetric distribution is acceptable in the special case where the regulator cares only about total surplus.[16] (Section 3.2.2 explains how, in a dynamic context, surplus can sometimes be returned to consumers over time.)

Countervailing incentives[17] In the foregoing setting, the firm's incentive is to exaggerate its cost in order to convince the regulator that more generous compensation is required to induce the firm to serve customers. This incentive to exaggerate private information may, in some circumstances, be tempered by a countervailing incentive to understate private information. To illustrate this effect, consider the following model.

Suppose everything is as specified above in the setting where realized costs are unobservable, with one exception. Suppose the level of fixed cost, F, is known only to the firm. It is common knowledge, though, that the firm's fixed cost is inversely related to its marginal cost, c.[18] In particular, it is common knowledge that when marginal cost is c_L, fixed cost is F_L, and that when marginal cost is c_H, fixed cost is $F_H < F_L$. Let $\Delta^F = F_L - F_H > 0$ denote the amount by which the firm's fixed cost increases as its marginal cost declines from c_H to c_L. As before, let $\Delta^c = c_H - c_L > 0$.

One might suspect that the regulator would suffer further when the firm is privately informed about both its fixed cost and its marginal cost of production rather than being privately informed only about the latter. This is not necessarily the case, though, as Proposition 2 reveals.

PROPOSITION 2. *When the firm is privately informed about both its fixed and its marginal cost*:

(i) *If* $\Delta^F \in [\Delta^c Q(c_H), \Delta^c Q(c_L)]$ *then the full-information outcome is feasible* (*and optimal*);

(ii) *If* $\Delta^F < \Delta^c Q(c_H)$ *then* $p_H > c_H$ *and* $p_L = c_L$;

(iii) *If* $\Delta^F > \Delta^c Q(c_L)$ *then* $p_L < c_L$ *and* $p_H = c_H$.

Part (i) of Proposition 2 considers a setting where the variation in fixed cost is of intermediate magnitude relative to the variation in variable cost when marginal-cost pricing is implemented. The usual incentive of the firm to exaggerate its marginal cost does

[16] This conclusion – derived here in an adverse selection setting – parallels the standard result that the full-information outcome can be achieved in a moral hazard setting when a risk-neutral agent is made the residual claimant for the social surplus. Risk neutrality in the moral hazard setting plays a role similar to the assumption here that distributional concerns are not present ($\alpha = 1$). The moral hazard problem is analyzed in Section 2.6 below.

[17] The following discussion is based on Lewis and Sappington (1989a). See Maggi and Rodriguez-Clare (1995) and Jullien (2000) for further analyses.

[18] If fixed costs increased as marginal costs increased, the firm would have additional incentive to exaggerate its marginal cost when it is privately informed about both fixed and marginal costs. Baron and Myerson (1982) show that the qualitative conclusions reported in Proposition 1 persist in this setting.

not arise at the full-information outcome in this setting. An exaggeration of marginal cost here amounts to an overstatement of variable cost by $\Delta^c Q(c_H)$. But it also constitutes an implicit understatement of fixed cost by Δ^F. Since Δ^F exceeds $\Delta^c Q(c_H)$, the firm would understate its true total operating cost if it exaggerated its marginal cost of production, and so will refrain from doing so. The firm also will have no incentive to understate its marginal cost at the full-information solution. Such an understatement amounts to a claim that variable costs are $\Delta^c Q(c_L)$ lower than they truly are. This understatement outweighs the associated exaggeration of fixed cost (Δ^F), and so will not be advantageous for the firm.

When the potential variation in fixed cost is either more pronounced or less pronounced than in part (i) of Proposition 2, the full-information outcome is no longer feasible. If the variation is less pronounced, then part (ii) of the proposition demonstrates that the qualitative distortions identified in Proposition 1 arise.[19] The prospect of understating fixed cost is no longer sufficient to eliminate the firm's incentive to exaggerate its marginal cost. Therefore, the regulator sets price above marginal cost when the firm claims to have high marginal cost in order to reduce the number of units of output ($Q(p_H)$) on which the firm can exercise its cost advantage.

In contrast, when the variation in fixed cost Δ^F exceeds $\Delta^c Q(c_L)$, the binding incentive problem for the regulator is to prevent the firm from exaggerating its fixed cost via understating its marginal cost. To mitigate the firm's incentive to understate c, part (iii) of Proposition 2 shows that the regulator sets p_L below c_L. Doing so increases beyond its full-information level the output the firm must produce in return for incremental compensation that is less than cost when the firm's marginal cost is high. Since the firm is not tempted to exaggerate its marginal cost (and thereby understate its fixed cost) in this setting, no pricing distortions arise when the high marginal cost is reported.

One implication of Proposition 2 is that the regulator may gain by *creating* countervailing incentives for the regulated firm. For instance, the regulator may mandate the adoption of technologies in which fixed costs vary inversely with variable costs. Alternatively, he may authorize expanded participation in unregulated markets the more lucrative the firm reports such participation to be (and thus the lower the firm admits its operating cost in the regulated market to be).[20]

Unknown scope for cost reduction[21] Now consider a setting where the regulator can observe the firm's marginal cost, but the firm's realized cost is affected by its (unobserved) cost-reducing effort, and the regulator is uncertain about the amount of effort required to achieve any given level of marginal cost.

[19] If $\Delta^F < \Delta^c Q(\hat{p}_H)$, where $\hat{p}_H = c_H + \frac{\phi}{1-\phi}(1-\alpha)\Delta^c$ is the optimal price for the high-cost firm identified in expression (5), then the price for the high-cost firm will be $p_H = \hat{p}_H$. Thus, for sufficiently small variation in fixed costs, the optimal pricing distortion is precisely the one identified by Baron and Myerson. The optimal distortion declines as Δ^F increases in the range $(\Delta^c Q(\hat{p}_H), \Delta^c Q(c_H))$.

[20] See Lewis and Sappington (1989a, 1989b, 1989c) for formal analyses of these possibilities.

[21] This is a simplified version of the model proposed in Laffont and Tirole (1986) and Laffont and Tirole (1993b, chs. 1 and 2). Also see Sappington (1982).

Suppose there are two types of firm. One (type L) can achieve low marginal cost via expending relatively low fixed cost. The other (type H) must incur greater fixed cost to achieve a given level of marginal cost. Formally, let $F_i(c)$ denote the fixed cost the type $i = L, H$ firm must incur to achieve marginal cost c. Each function $F_i(\cdot)$ is decreasing and convex, where $F_H(c) > F_L(c)$ and where $[F_H(c) - F_L(c)]$ is a decreasing function of c. The regulator cannot observe the firm's type, and views it as a random variable that takes on the value L with probability $\phi \in (0, 1)$ and H with probability $1 - \phi$. As noted, the regulator can observe the firm's realized marginal cost c in the present setting, but cannot observe the associated realization of the fixed cost $F_i(c)$.

Because realized marginal cost is observable, the regulator has three policy instruments at his disposal. He can specify a unit price (p) for the firm's product, a transfer payment (T) from consumers to the firm, and a realized level of marginal cost (c). Therefore, for each $i = L, H$ the regulator announces that he will authorize price p_i and transfer payment T_i when the firm claims to be of type i, provided marginal cost c_i is observed. The equilibrium rent of the type i firm, R_i, is then

$$R_i = Q(p_i)(p_i - c_i) - F_i(c_i) + T_i. \tag{7}$$

As in inequality (4) above, the constraint that the low-cost firm does not claim to have high cost is

$$R_L \geqslant R_H + F_H(c_H) - F_L(c_H). \tag{8}$$

Net consumer surplus in state i is $v(p_i) - T_i$. Using equality (7), this net consumer surplus can be written as

$$v(p_i) + Q(p_i)(p_i - c_i) - F_i(c_i) - R_i. \tag{9}$$

Notice that the regulator's choice of prices $\{p_L, p_H\}$ does not affect the incentive constraint (8), given the choice of rents $\{R_L, R_H\}$. Therefore, price will be set equal to realized marginal cost (i.e., $p_i = c_i$) in order to maximize total surplus in (9). This conclusion reflects Laffont and Tirole's "incentive-pricing dichotomy": prices (often) should be employed solely to attain allocative efficiency, while rents should be employed to motivate the firm to produce at low cost.[22] This dichotomy represents a key difference between Baron and Myerson's (1982) model in which costs are exogenous and unobservable and Laffont and Tirole's (1986) model in which costs are endogenous and observable. In the former model, price levels are distorted in order to reduce the firm's rent. In the latter model, price levels are not distorted given the induced costs, but cost distortions are implemented to limit the firm's rent.

If the regulator knew the firm's type, he would also require the efficient marginal cost, which is the cost that maximizes total surplus $\{v(c) - F_i(c)\}$. However, the full-information outcome is not feasible when the regulator does not share the firm's

[22] For further analysis of the incentive-pricing dichotomy, including a discussion of conditions under which the dichotomy does not hold, see Laffont and Tirole (1993b, Sections 2.3 and 3.6).

knowledge of its technology. To limit the type-L firm's rent, the regulator inflates the type-H firm's marginal cost above the full-information level, as reported in Proposition 3.

PROPOSITION 3. *When the firm's marginal cost is observable but endogenous, the optimal regulatory policy has the following features*:

$$p_L = c_L; \qquad p_H = c_H; \tag{10}$$

$$Q(c_L) + F_L'(c_L) = 0; \tag{11}$$

$$Q(c_H) + F_H'(c_H) = -\frac{\phi}{1-\phi}(1-\alpha)\big(F_H'(c_H) - F_L'(c_H)\big) > 0, \tag{12}$$

$$R_L = F_H(c_H) - F_L(c_H) > 0; \qquad R_H = 0. \tag{13}$$

Expression (11) indicates that the type-L firm will be induced to operate with the cost-minimizing technology. In contrast, expression (12) reveals that the type-H firm will produce with inefficiently high marginal cost. This high marginal cost limits the rent that accrues to the type-L firm, which, from inequality (8), decreases as c_H increases. As expression (12) reveals, the optimal distortion in c_H is more pronounced the more likely is the firm to have low cost (i.e., the larger is $\phi/(1-\phi)$) and the more the regulator cares about minimizing rents (i.e., the smaller is α). The marginal cost implemented by the low-cost firm is not distorted because the high-cost firm is not tempted to misrepresent its type.[23,24]

2.3.2. Asymmetric demand information

The analysis to this point has assumed that the demand function facing the firm is common knowledge. In practice, regulated firms often have privileged information about consumer demand. To assess the impact of asymmetric knowledge of this kind, consider the following simple model.[25]

The firm's cost function, $C(\cdot)$, is common knowledge, but consumer demand can take one of two forms: the demand function is $Q_L(\cdot)$ with probability ϕ and $Q_H(\cdot)$

[23] The regulator may implement other distortions when he has additional policy instruments at his disposal. For example, the regulator may require the firm to employ more than the cost-minimizing level of capital when additional capital reduces the sensitivity of realized costs to the firm's unobserved innate cost. By reducing this sensitivity, the regulator is able to limit the rents that the firm commands from its privileged knowledge of its innate costs. See Sappington (1983) and Besanko (1985), for example.

[24] Extending the analysis of Guesnerie and Laffont (1984), Laffont and Rochet (1998) examine how risk aversion on the part of the regulated firm affects the optimal regulatory policy in a setting where the firm's realized marginal cost is observable and endogenous. The authors show that risk aversion introduces more pronounced cost distortions, reduces the rent of the firm, and may render realized marginal cost insensitive to the firm's innate capabilities over some ranges of capability.

[25] The following discussion is based on Lewis and Sappington (1988a).

with probability $1 - \phi$, where $Q_H(p) > Q_L(p)$ for all prices p. The firm knows the demand function it faces from the outset of its relationship with the regulator. The regulator never observes the prevailing demand function. Furthermore, the regulator never observes realized cost or realized demand.[26] The firm is required to serve all customer demand and will operate as long as it receives non-negative profit from doing so.

As in the setting with countervailing incentives, the regulator's limited information need not be constraining here. To see why in the simplest case, suppose the firm's cost function is affine, i.e., $C(q) = cq + F$, where q is the number of units of output produced by the firm. In this case, the regulator can instruct the firm to sell its product at price equal to marginal cost in return for a transfer payment equal to F. Doing so ensures marginal-cost pricing and zero rent for the firm for both demand realizations, which is the full-information outcome. When marginal cost is constant, the full-information pricing policy (i.e., $p = c$) is common knowledge because it depends only on the firm's (known) marginal cost of production.[27]

More surprisingly, Proposition 4 states that the regulator can also ensure the full-information outcome if marginal cost increases with output.

PROPOSITION 4. *In the setting where the firm is privately informed about demand*:

(i) *If* $C''(q) \geqslant 0$, *the full-information outcome is feasible* (*and optimal*);

(ii) *If* $C''(q) < 0$, *the regulator often*[28] *sets a single price and transfer payment for all demand realizations.*

When marginal cost increases with output, the full-information price for the firm's product p increases with demand, and the transfer payment to the firm T declines with demand. The higher price reflects the higher marginal cost of production that accompanies increased output. The reduction in T just offsets the higher variable profit the firm secures from the higher p. Since the reduction in T exactly offsets the increase in variable profit when demand is high, it more than offsets any increase in variable profit from a higher p when demand is low. Therefore, the firm has no incentive to exaggerate demand. When demand is truly low, the reduction in T that results when demand is exaggerated more than offsets the extra profit from the higher p that is authorized. Similarly, the firm has no incentive to understate demand when the regulator offers the firm two choices that constitute the full-information outcome. The understatement of demand calls forth a price reduction that reduces the firm's profit by more than the

[26] If he could observe realized costs or demand, the regulator would be able to infer the firm's private information since he knows the functional forms of $C(\cdot)$ and $Q_i(\cdot)$.

[27] This discussion assumes that production is known to be desirable for all states of demand.

[28] The precise meaning of "often" is made clear in Section 2.3.3. To illustrate, pooling is optimal when the two demand functions differ by an additive constant and $C(\cdot)$ satisfies standard regularity conditions.

corresponding increase in the transfer payment it receives.[29] In sum, part (i) of Proposition 4 states that the full-information outcome is feasible in this setting.[30]

Part (ii) of Proposition 4 shows that the same is not true when marginal cost declines with output. In this case, the optimal price p declines as demand increases in the full-information outcome.[31] In contrast, in many reasonable cases, the induced price p cannot decline as demand increases when the firm alone knows the realization of demand. A substantial increase in the transfer payment T would be required to compensate the firm for the decline in variable profit that results from a lower p when demand is high. This increase in T more than compensates the firm for the corresponding reduction in variable profit when demand is low. Consequently, the firm cannot be induced to set a price that declines as demand increases. When feasible prices increase with demand while full-information prices decline with demand, the regulator is unable to induce the firm to employ its private knowledge of demand to benefit consumers. Instead, he chooses a single unit price and transfer payment to maximize expected welfare. Thus, when the firm's cost function is concave, it is too costly from a social point of view to make use of the firm's private information about demand.[32]

Notice that in the present setting where there is no deadweight loss involved in raising tax revenue, the relevant full-information benchmark is marginal-cost pricing. As noted in Section 2.1, when a transfer payment to the firm imposes a deadweight loss on society, Ramsey prices become the relevant full-information benchmark. Since the implementation of Ramsey prices requires knowledge of consumer demand, the regulator will generally be unable to implement the full-information outcome when he is ignorant about consumer demand, even when the firm's cost function is known to be convex. Consequently, the qualitative conclusion drawn in Proposition 4 does not extend to the

[29] Lewis and Sappington (1988a) show that the firm has no strict incentive to understate demand in this setting even if it can ration customers with impunity. The authors also show that the arguments presented here are valid regardless of the number of possible states of demand. Riordan (1984) provides a corresponding analysis for the case where the firm's marginal cost is constant up to an endogenous capacity level. Lewis and Sappington (1992) show that part (i) of Proposition 4 continues to hold when the regulated firm chooses the level of observable and contractible quality it supplies.

[30] Biglaiser and Ma (1995) analyze a setting in which a regulated firm produces with constant marginal cost and is privately informed about both the demand for its product and the demand for the (differentiated) product of its unregulated rival. The authors show that when the regulator's restricted set of instruments must serve both to limit the rents of the regulated firm and to limit the welfare losses that result from the rival's market power, the optimal regulatory policy under asymmetric information differs from the corresponding policy under complete information. Therefore, part (i) of Proposition 4 does not always hold when the regulated firm faces an unregulated rival with market power.

[31] This will be the case when the marginal cost curve is less steeply sloped than the inverse demand curve, and so the regulator's problem is concave and there exists a unique welfare-maximizing price that equals marginal cost in each state.

[32] A similar finding emerges in Section 2.5, where the regulator's intertemporal commitment powers are limited. In that setting, it can be too costly to induce the low-cost firm to reveal its superior capabilities, because the firm fears the regulator will expropriate all future rent. Consequently, the regulator may optimally implement some pooling in order to remain ignorant about the firm's true capabilities.

setting where transfer payments to the firm are socially costly. In contrast, the qualitative conclusions drawn in Propositions 1, 2 and 3 persist in the alternative setting when transfer payments are socially costly, provided the full-information prices are Ramsey prices rather than marginal-cost prices.

2.3.3. A unified analysis

The foregoing analyses reveal that the qualitative properties of optimal regulatory policies can vary substantially according to the nature of the firm's private information and its technology. Optimal regulated prices can be set above, below, or at the level of marginal cost, and the full-information outcome may or may not be feasible, depending on whether the firm is privately informed about the demand function it faces, its variable production costs, or both its variable and its fixed costs of production. The purpose of this subsection is to explain how these seemingly disparate findings all emerge from a single, unified framework.[33] This section also provides a sketch of the proofs of the propositions presented above. Consequently, this section is somewhat more technical than most. The less technically-oriented reader can skip this section without compromising understanding of subsequent discussions.

This unifying framework has the following features. The firm's private information takes on one of two possible values, which will be referred to as state L or state H. The probability of state L is $\phi \in (0, 1)$ and the probability of state H is $1 - \phi$. The firm's operating profit in state i when it charges unit price p for its product is $\pi_i(p)$. The firm's equilibrium rent in state i is $R_i = \pi_i(p_i) + T_i$, where p_i is the price for the firm's product and T_i is the transfer payment from the regulator to the firm in state i.

The difference in the firm's operating profit at price p in state H versus state L will be denoted $\Delta^\pi(p)$. For most of the following analysis, this difference is assumed to increase with p. Formally,

$$\Delta^\pi(p) \equiv \pi_H(p) - \pi_L(p) \quad \text{and} \quad \frac{\mathrm{d}}{\mathrm{d}p}\Delta^\pi(p) > 0. \tag{14}$$

The "increasing difference" property in expression (14) reflects the standard single crossing property.[34] Its role, as will be shown below, is to guarantee that the equilibrium price in state H is higher than in state L.

The regulator seeks to maximize the expected value of a weighted average of consumer surplus and rent. The social cost of public funds is assumed to be zero. Consumer surplus in state i given price p is the surplus (denoted $v_i(p)$) from consuming

[33] This material is taken from Armstrong and Sappington (2004). Guesnerie and Laffont (1984) and Caillaud et al. (1988) provide earlier unifying analyses of adverse selection models in the case where private information is a continuously distributed variable. Although the qualitative features of the solutions to continuous and discrete adverse selection problems are often similar, the analytic techniques employed to solve the two kinds of problems differ significantly.

[34] The single crossing property holds when the firm's marginal rate of substitution of price for transfer payment varies monotonically with the underlying state. See Cooper (1984) for details.

the product at price p, minus the transfer, T_i, to the firm. Written in terms of rent $R_i = \pi_i(p_i) + T_i$, this weighted average of consumer surplus and rent in state i is

$$S_i + \alpha R_i = v_i(p_i) - T_i + \alpha\bigl(\pi_i(p_i) + T_i\bigr) = w_i(p_i) - (1-\alpha)R_i. \tag{15}$$

Here, $w_i(p) \equiv v_i(p) + \pi_i(p)$ denotes total unweighted surplus in state i when price p is charged, and $\alpha \in [0, 1]$ is the relative weight placed on rent in social welfare. Therefore, expected welfare is

$$W = \phi\bigl\{w_L(p_L) - (1-\alpha)R_L\bigr\} + (1-\phi)\bigl\{w_H(p_H) - (1-\alpha)R_H\bigr\}. \tag{16}$$

The type i firm will agree to produce according to the specified contract only if it receives a non-negative rent. Consequently, the regulator faces the two participation constraints

$$R_i \geqslant 0 \quad \text{for } i = L, H. \tag{17}$$

If the regulator knew that state i was the prevailing state, he would implement the price p_i^* that maximizes $w_i(\cdot)$ while ensuring $R_i = 0$. This is the full-information benchmark, and involves marginal-cost pricing. If the regulator does not know the prevailing state, he must ensure that contracts are such that each type of firm finds it in its interest to choose the correct contract. Therefore, as in expressions (4) and (8) above, the regulator must ensure that the following incentive compatibility constraints are satisfied:

$$R_L \geqslant R_H - \Delta^\pi(p_H), \tag{18}$$

$$R_H \geqslant R_L + \Delta^\pi(p_L). \tag{19}$$

Adding inequalities (18) and (19) implies

$$\Delta^\pi(p_H) \geqslant \Delta^\pi(p_L). \tag{20}$$

The increasing difference assumption in expression (14) and inequality (20) together imply the equilibrium price must be weakly higher in state H than in state L in any incentive-compatible regulatory policy, i.e.,

$$p_H \geqslant p_L. \tag{21}$$

The following conclusion aids in understanding the solution to the regulator's problem.

LEMMA 1. *If the incentive compatibility constraint for the type i firm does not bind at the optimum, then the price for the other type of firm is not distorted, i.e.,* $p_j = p_j^*$.[35]

[35] The surplus functions $w_i(p_i)$ are assumed to be single-peaked.

To understand this result, suppose the incentive compatibility constraint for the type-H firm, inequality (19), does not bind at the optimum. Then, holding R_L constant – which implies that neither the participation constraint nor the incentive compatibility constraint for the type-L firm is affected – the price p_L can be changed (in either direction) without violating (19). If a small change in p_L does not increase welfare $w_L(p_L)$ in (16), then p_L must (locally) maximize $w_L(\cdot)$, which proves Lemma 1.

Now consider some special cases of this general framework.

When is the full-information outcome feasible? Recall that in the full-information outcome, the type-i firm sets price p_i^* and receives zero rent.[36] The incentive constraints (18) and (19) imply that this full-information outcome is attainable when the regulator does not observe the state if and only if

$$\Delta^\pi\left(p_H^*\right) \geqslant 0 \geqslant \Delta^\pi\left(p_L^*\right). \tag{22}$$

The pair of inequalities in (22) imply that the full-information outcome will not be feasible if the firm's operating profit $\pi(p)$ is systematically higher in one state than the other (as when the firm is privately informed only about its marginal cost of production, for example). If the full-information outcome is to be attainable, the profit functions $\pi_H(\cdot)$ and $\pi_L(\cdot)$ must cross: operating profit must be higher in state H than in state L at the full-information price p_H^*, and operating profit must be lower in state H than in state L at the full-information price p_L^*.

Recall from part (i) of Proposition 4 that the full-information outcome is feasible in the setting where the firm's convex cost function $C(\cdot)$ is common knowledge but the firm is privately informed about the demand function it faces. In this context, demand is either high, $Q_H(\cdot)$, or low, $Q_L(\cdot)$, and the profit function in state i is $\pi_i(p) = pQ_i(p) - C(Q_i(p))$. To see why the full-information outcome is feasible in this case, let $q_i^* \equiv Q_i(p_i^*)$ denote the firm's output in state i in the full-information outcome. Since $C(\cdot)$ is convex:

$$C\left(Q_i\left(p_j^*\right)\right) \geqslant C\left(q_j^*\right) + C'\left(q_j^*\right)\left(Q_i\left(p_j^*\right) - q_j^*\right). \tag{23}$$

To show that inequality (22) is satisfied when prices are equal to marginal costs, notice that

$$\begin{aligned}
\pi_i\left(p_j^*\right) &= p_j^* Q_i\left(p_j^*\right) - C\left(Q_i\left(p_j^*\right)\right) \\
&\leqslant p_j^* Q_i\left(p_j^*\right) - \left\{C\left(q_j^*\right) + C'\left(q_j^*\right)\left(Q_i\left(p_j^*\right) - q_j^*\right)\right\} \\
&= p_j^* q_j^* - C\left(q_j^*\right) \\
&= \pi_j\left(p_j^*\right).
\end{aligned} \tag{24}$$

[36] The single-crossing condition $\frac{\mathrm{d}}{\mathrm{d}p}\Delta^\pi > 0$ is not needed for the analysis in this section.

The inequality in expression (24) follows from inequality (23). The second equality in expression (24) holds because $p_j^* = C'(q_j^*)$. Consequently, condition (22) is satisfied, and the regulator can implement the full-information outcome.

Part (i) of Proposition 2 indicates that the full-information outcome is also feasible in the setting where the demand function facing the firm is common knowledge, the firm is privately informed about its constant marginal c_i and fixed costs F_i of production, and the variation in fixed cost is intermediate in magnitude. To prove this conclusion, we need to determine when the inequalities in (22) are satisfied. Since $\pi_i(p) = (p - c_i)Q(p) - F_i$ in this setting, it follows that

$$\Delta^\pi(p) = \Delta^F - \Delta^c Q(p). \tag{25}$$

Therefore, since full-information prices are $p_i^* \equiv c_i$, expression (25) implies that the inequalities (22) will be satisfied if and only if $\Delta^c Q(c_L) \geqslant \Delta^F \geqslant \Delta^c Q(c_H)$, as indicated in Proposition 2.

Price distortions with separation Suppose that profit is higher in state L than in state H for all prices, so that $\Delta^\pi(p) < 0$. In this case, only the type-H firm's participation constraint in (17) will be relevant, and this firm will optimally be afforded no rent.[37] In this case, the incentive compatibility constraints (18)–(19) become

$$-\Delta^\pi(p_L) \geqslant R_L \geqslant -\Delta^\pi(p_H).$$

Again, since rent is costly, it is only the lower bound on R_L that is relevant, i.e., only the incentive compatibility constraint (18) is relevant.

Therefore, expression (16) reduces to

$$W = \phi\big\{w_L(p_L) + (1-\alpha)\Delta^\pi(p_H)\big\} + (1-\phi)w_H(p_H). \tag{26}$$

Notice that this expression for W incorporates both the type-H firm's participation constraint and the type-L firm's incentive compatibility constraint.

Maximizing expression (26) with respect to p_L and p_H implies:

$$p_L = p_L^* \quad \text{and } p_H \text{ maximizes} \quad w_H(p) + \frac{\phi}{1-\phi}(1-\alpha)\Delta^\pi(p). \tag{27}$$

Since $\Delta^\pi(p)$ increases with p, expression (27) implies that $p_H \geqslant p_H^*$. When full-information prices are ordered as in inequality (21), it follows that $p_H \geqslant p_H^* \geqslant p_L^* = p_L$, and therefore the monotonicity condition (21) is satisfied. Therefore, (27) provides the solution to the regulator's problem. In particular, the regulator will induce the firm to set different prices in different states, and the type H firm's price will be distorted above the full-information level, p_H^*. This distortion is greater the more costly are rents (the lower is α) and the more likely is state L (the higher is ϕ).

[37] Since rents are costly in expression (16) and the incentive compatibility constraints (18)–(19) depend only on the difference between the rents, at least one participation constraint must bind at the optimum.

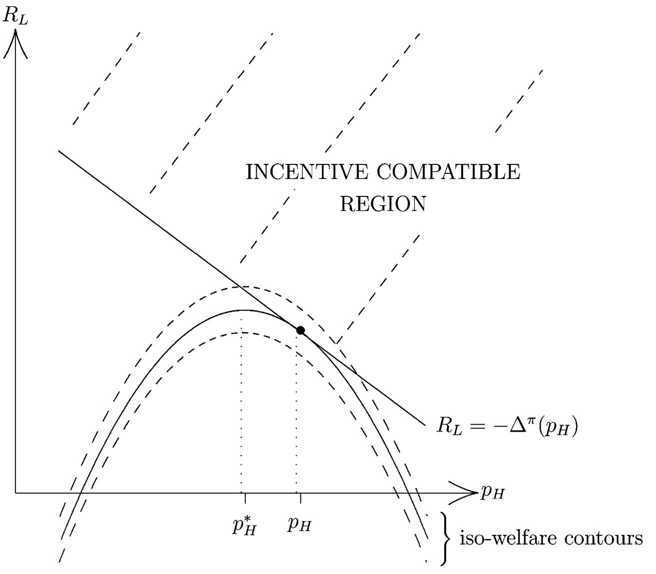

Figure 27.1. Price distortions with separation.

This analysis is presented in Figure 27.1, which depicts outcomes in terms of p_H and R_L. (The remaining choice variables are R_H, which is set equal to zero at the optimum, and p_L, which is set equal to p_L^* at the optimum.) Here, the incentive compatible region is the set of points $R_L \geqslant -\Delta^\pi(p_H)$, and the regulator must limit himself to a contract that lies within this set. Expression (16) shows that welfare contours in (p_H, R_L) space take the form $R_L = (\frac{1}{1-\alpha})(\frac{1-\phi}{\phi})w_H(p_H) + k_o$, where k_o is a constant. Each of these contours is maximized at $p_H = p_H^*$, as shown in Figure 27.1. (Lower contours indicate higher welfare.) Therefore, the optimum is where a welfare contour just meets the incentive compatible region, which necessarily involves a price p_H greater than p_H^*. Increasing α, so that distributional concerns are less pronounced, or reducing ϕ, so that the high-cost state is more likely, steepens the welfare contours, and so brings the optimal choice of price p_H closer to the full-information price p_H^*.

These qualitative features characterize the optimal regulatory policy in many settings, including the Baron–Myerson setting of Proposition 1 where the firm is privately informed about its marginal cost of production. In this setting, $\pi_i(p) = Q(p)(p - c_i) - F$, and so $\Delta^\pi(p) = -\Delta^c Q(p)$. Therefore, expression (27) implies that the optimal price for the high-cost firm is as given in expression (5) of Proposition 1. Notice that in the context of the Baron–Myerson model, welfare in expression (26) can be written as

$$W = \phi\{v(p_L) + Q(p_L)(p_L - c_L)\} + (1-\phi)\{v(p_H) + Q(p_H)(p_H - \hat{c}_H)\}, \tag{28}$$

where

$$\hat{c}_H = c_H + \frac{\phi}{1-\phi}(1-\alpha)\Delta^c. \tag{29}$$

Here, $\hat{c}_H$ is simply p_H in expression (5).[38] Expression (28) reveals that expected welfare is the same in the following two situations: (a) the firm has private information about its marginal cost, where this cost is either c_L or c_H; and (b) the regulator can observe the firm's marginal cost, where this cost is either c_L or $\hat{c}_H$. (Of course, the firm is better off under situation (a).) Thus, the effect of private information on welfare in this setting is exactly the effect of inflating the cost of the high-cost firm according to formula (29) in a setting with no information asymmetry. Under this interpretation, the prices in expression (5) are simply marginal-cost prices, where the "costs" are increased above actual costs to account for socially undesirable rents caused by asymmetric information.

Similar conclusions emerge in the Laffont–Tirole model with observed but endogenous marginal cost, as in Proposition 3. Here, once it is noted that price is optimally set equal to the realized and observed marginal cost ($p_i \equiv c_i$), the problem fits the current framework precisely. Specifically, $\pi_i(p_i) = -F_i(p_i)$, and so $\Delta^\pi(p) = F_L(p) - F_H(p) < 0$, which is assumed to increase with p. Also, $w_i(p) = v(p) - F_i(p)$. Therefore, expression (27) yields expression (12) in Proposition 3.

Pooling It remains to illustrate why the regulator might sometimes implement the same contract in both states of the world. As suggested in the discussion after Proposition 4, pooling (i.e., $p_H = p_L$) may be optimal if $p_L^* > p_H^*$, so that prices in the full-information outcome do not satisfy the necessary condition for incentive compatibility, inequality (21).

To illustrate this observation, consider the setting where the firm's strictly concave cost function is common knowledge and the firm is privately informed about its demand function. Because marginal cost declines with output in this setting, the full-information prices satisfy $p_L^* > p_H^*$. Whether the single-crossing condition (14) is satisfied depends on joint effects of the cost function and the variation in demand. One can show, for instance, that if $Q_H(\cdot) \equiv Q_L(\cdot) + k_1$ where k_1 is a constant, then the single-crossing condition will be satisfied if the cost function satisfies certain standard regularity conditions.

To see that pooling is optimal in this setting, suppose to the contrary that separation ($p_L \neq p_H$) is implemented at the optimum. Then it is readily verified that exactly one of the incentive compatibility constraints (18) or (19) binds, and so, from Lemma 1, the full-information price p_i^* is implemented in one state.[39] Suppose state L is the relevant

[38] The adjusted cost $\hat{c}_H$ is often referred to as the "virtual cost".

[39] If the single-crossing condition holds, both incentive constraints can only bind when $p_L = p_H$. If neither constraint binds, then $p_L = p_L^*$ and $p_H = p_H^*$, which cannot be incentive compatible when $p_L^* > p_H^*$.

state, so that $p_L = p_L^*$ and (18) binds:

$$R_L = R_H - \Delta^\pi(p_H). \tag{30}$$

An analysis analogous to that which underlies expressions (23) and (24) reveals that $\pi_H(p_L^*) \geqslant \pi_L(p_L^*)$ when $C(\cdot)$ is concave, i.e., $\Delta^\pi(p_L^*) \geqslant 0$. Since inequality (21) requires $p_H > p_L^*$ and since $\Delta^\pi(\cdot)$ is an increasing function of p, it follows that $\Delta^\pi(p_H) > 0$. Since at least one participation constraint (17) binds, expression (30) implies

$$R_L = 0; \qquad R_H = \Delta^\pi(p_H). \tag{31}$$

Therefore, expected welfare in expression (16) simplifies to

$$W = \phi w_L\big(p_L^*\big) + (1-\phi)\big\{w_H(p_H) - (1-\alpha)\Delta^\pi(p_H)\big\}. \tag{32}$$

Since $p_H > p_H^*$, it follows that the $\{\cdot\}$ term in expression (32) is decreasing in p_H if $w_H(\cdot)$ is single-peaked in price. Since a small reduction in p_H does not violate any participation or incentive compatibility constraint and increases the value of the regulator's objective function, the candidate prices $\{p_L^*, p_H\}$ cannot be optimal. A similar argument holds if inequality (19) is the binding incentive constraint. Therefore, by contradiction, $p_L = p_H$ in the solution to the regulator's problem.[40]

Notice that, in contrast to the pricing distortions discussed above (e.g., in expression (27)), pooling is not implemented here to reduce the firm's rent. Even if the regulator valued rent and consumer surplus equally (so $\alpha = 1$), pooling would still be optimal in this setting. Pooling arises here because of the severe constraints imposed by incentive compatibility.

2.4. Extensions to the basic model

The analysis to this point has been restrictive because: (i) the regulator had no opportunity to obtain better information about the prevailing state; and (ii) the regulator was uninformed about only a single "piece" of relevant information. In this section, two alternative information structures are considered. First, the regulator is allowed to obtain some imperfect information about the realized state, perhaps through an audit. Two distinct settings are examined in this regard: one where the regulator always acts in the interests of society, and one where the firm may bribe the regulator to conceal potentially damaging information. The latter setting permits an analysis of how the danger of regulatory capture affects the optimal design of regulation. Second, the firm is endowed with superior information about more than one aspect of its environment. We illustrate each of these extensions by means of natural variants of the Baron–Myerson model discussed in Section 2.3.1.

[40] Laffont and Martimort (2002, Section 2.10.2) provide further discussion of when pooling will arise in models of this sort.

2.4.1. Partially informed regulator: the use of audits

First consider the setting where the firm is privately informed about its exogenous constant marginal cost of production ($c \in \{c_L, c_H\}$). Suppose that an imperfect public signal $s \in \{s_L, s_H\}$ of the firm's cost is available, which is realized after contracts have been signed. This signal is "hard" information in the sense that (legally enforceable) contracts can be written based on this information. This signal might be interpreted as the output of an audit of the firm's cost, for example. Specifically, let ϕ_i denote the probability that low signal s_L is observed when the firm's marginal cost is c_i for $i = L, H$. To capture the fact that the low signal is likely to be associated with low cost, assume $\phi_L > \phi_H$.[41]

Absent bounds on the rewards or penalties that can be imposed on the risk-neutral firm, the regulator can ensure marginal-cost pricing without ceding any rent to the firm in this setting. He can do so by conditioning the transfer payment to the firm on the firm's report of its cost and on the subsequent realization of the public signal. Specifically, let T_{ij} be the regulator's transfer payment to the firm when the firm claims its cost is c_i and when the realized signal turns out to be s_j. If the firm claims to have a high cost, it is permitted to charge the (high) unit price, $p_H = c_H$. In addition, it receives a generous transfer payment when the signal (s_H) suggests that its cost is truly high, but is penalized when the signal (s_L) suggests otherwise. These transfer payments can be structured to provide an expected transfer which just covers the firm's fixed cost F when its marginal cost is indeed c_H, i.e.,

$$\phi_H T_{HL} + (1 - \phi_H) T_{HH} = F. \tag{33}$$

At the same time, the payments can be structured to provide an expected return to the low-cost firm that is sufficiently far below F that they eliminate any rent the low-cost firm might otherwise anticipate from being able to set the high price ($p_H = c_H$), i.e.,

$$\phi_L T_{HL} + (1 - \phi_L) T_{HH} \ll F. \tag{34}$$

The transfers T_{HL} and T_{HH} can always be set to satisfy equality (33) and inequality (34) except in the case where the signal is entirely uninformative ($\phi_L = \phi_H$). The low-cost firm can simply be offered its (deterministic) full-information contract, with price p_L equal to marginal cost c_L and transfer $T_{LH} = T_{LL}$ equal to the fixed cost F. This pair of contracts secures the full-information outcome. Thus, even an imprecise monitor of the firm's private cost information can constitute a powerful regulatory instrument when feasible payments to the firm are not restricted and when the firm is risk neutral.[42]

[41] Another way to model this audit would be to suppose the regulator observes the true cost with some exogenous probability (and otherwise observes "nothing"). This alternative specification yields the same insights. A form of this alternative specification is explored in Section 2.4.2, which discusses regulatory capture.

[42] This insight will play an important role in the discussion of yardstick competition in Section 4.1.2, where, instead of an audit, the signal about one firm's cost is obtained from the report of a second firm with correlated costs. Crémer and McLean (1985), Riordan and Sappington (1988) and Caillaud, Guesnerie and Rey (1992) provide corresponding conclusions in more general settings.

When the maximum penalty that can be imposed on the firm ex post is sufficiently small in this setting, the low-cost firm will continue to earn rent.[43] To limit these rents, the regulator will implement the qualitative pricing distortions identified in Proposition 1.[44] Similar rents and pricing distortions will also arise if risk aversion on the part of the firm makes the use of large, stochastic variations in transfer payments to the firm prohibitively costly.[45]

If the regulator has to incur a cost to receive the signal from an audit, the regulator will have to decide when to purchase the signal.[46] If there were no constraints on the size of feasible punishments, the full-information outcome could be approximated arbitrarily closely. The regulator could undertake a costly audit with very small probability and punish the firm very severely if the signal contradicts the firm's report. In contrast, when the magnitude of feasible punishments is limited, the full-information outcome can no longer be approximated. Instead, the regulator will base his decision about when to purchase the signal on the firm's report. If the firm announces it has low cost, then no audit is commissioned, and price is set at the full-information level. In contrast, if the firm claims to have high cost, the regulator commissions an audit with a specified probability.[47] The frequency of this audit is determined by balancing the costs of auditing with the benefits of improved information.

2.4.2. Partially informed regulator: regulatory capture

The discussion in this section relaxes the assumption that the regulator automatically acts in the interests of society.[48] For simplicity, consider the other extreme in which the regulator aims simply to maximize his personal income. This income may arise from two sources. First, the firm may attempt to "bribe" the regulator to conceal information that is damaging to the firm. Second, and in response to this threat of corruption, the regulator himself may operate under an incentive scheme, which rewards him when he

[43] See Baron and Besanko (1984a), Demougin and Garvie (1991), and Gary-Bobo and Spiegel (2006), for example.

[44] See Baron and Besanko (1984b).

[45] See Baron and Besanko (1987b).

[46] See Baron and Besanko (1984b) and Laffont and Martimort (2002, Section 3.6).

[47] The importance of the regulator's presumed ability to commit to an auditing policy is apparent. See Khalil (1997) for an analysis of the setting where the regulator cannot commit to an auditing strategy.

[48] This discussion is based on Laffont and Tirole (1991b) and Laffont and Tirole (1993b, ch. 11). To our knowledge, Tirole (1986a) provides the first analysis of these three-tier models with collusion. Demski and Sappington (1987) also study a three-tier model, but their focus is inducing the regulator to monitor the firm. (The regulator incurs a private cost if he undertakes an audit, but the firm does not attempt to influence the regulator's behavior.) Spiller (1990) provides a moral hazard model in which, by expending unobservable effort, the regulator can affect the probability of the firm's price being high or low. In this model, the firm and the political principal try to influence the regulator's choice of effort by offering incentives based on the realized price.

reveals this damaging information. This incentive scheme is designed by a "political principal", who might be viewed as the (benevolent) government, for example.[49]

To be specific, suppose the firm's marginal cost is either c_L or c_H. The probability of the low-cost (c_L) realization is ϕ. Also, suppose that conditional on the firm's cost realization being low, the regulator has an exogenous probability ζ of being informed that the cost is indeed low. Conditional on a high-cost realization, the regulator has no chance of being informed.[50] The probability that the regulator is informed (which implies that the firm has low cost) is $\psi = \phi\zeta$. The probability that the regulator is uninformed is $1 - \psi$. The probability of the cost being low, conditional on the regulator being uninformed, therefore, is

$$\phi^U = \frac{\phi(1-\zeta)}{1-\phi\zeta} < \phi.$$

The information obtained by the regulator is "hard" in the sense that revelation of the regulator's private signal that cost is low proves beyond doubt that the firm has low cost. Therefore, when the regulator admits to being informed, the (low-cost) firm is regulated with symmetric information and so the firm receives no rent. However, if the regulator claims to be uninformed, the principal is unable to confirm this is in fact the case. The principal is unable to determine whether the firm and regulator have successfully colluded and the regulator is concealing the damaging information he has actually obtained.

Suppose the regulator must be paid at least zero by the principal in every state.[51] Also suppose the principal pays the regulator an extra amount s when the regulator admits to being informed. Assume for now that the principal induces the regulator to reveal his information whenever he is informed, i.e., that the principal implements a "collusion-proof" mechanism. In this case, when the regulator announces he is uninformed, the probability that the firm has low cost is ϕ^U. This probability becomes the relevant probability of having a low-cost realization when calculating the optimal regulatory policy in this case.

49 An alternative formulation would have the regulator commission an auditor to gather information about the firm. The firm might then try to bribe the auditor to conceal detrimental information from the regulator.

50 Laffont and Tirole (1993b, ch. 11) model the information structure more symmetrically by assuming the regulator is informed about the true cost with probability ζ whether the cost is high or low. However, when the regulator learns the cost is high, the firm has no interest in persuading him to conceal this information. Since the possibility that the regulator might learn that cost is high plays no significant role in this analysis of capture (but complicates the notation) we assume the regulator can obtain information only about a low cost realization.

51 The ex post nature of this participation constraint for the regulator is important. If the regulator were risk neutral and cared only about expected income, he could be induced to reveal his information to the political principal at no cost. (This could be done by offering the regulator a high reward when he revealed information and a high penalty when he claimed to be uninformed, with these two payments set to ensure the regulator zero expected rent, in a manner similar to that described in Section 2.4.1 above.) In addition, by normalizing the regulator's required income to zero, we introduce the implicit assumption that the regulator is indispensable for regulation, and the political principal cannot do without his services and cannot avoid paying him his reservation wage.

Suppose it costs the firm \$$(1+\theta)$ to increase the income of the regulator by \$1. The deadweight loss θ involved in increasing the regulator's income may reflect legal restrictions designed to limit the ability of regulated firms to influence regulators unduly, for example. These restrictions include prohibitions on direct bribery of government officials. Despite such prohibitions, a firm may find (costly) ways to convince the regulator of the merits of making decisions that benefit the firm. For instance, the firm may provide lucrative employment opportunities for selected regulators or agree to charge a low price for a politically-sensitive service in return for higher prices on other services. For simplicity, we model these indirect ways of influencing the regulator's decision as an extra marginal cost θ the firm incurs in delivering income to the regulator. For expositional ease, we will speak of the firm as "bribing" the regulator, even though explicit bribery may not actually be undertaken.[52]

It is clear from Proposition 1 that the low-cost firm will optimally be induced to set a price equal to its cost. Suppose that when the regulator is uninformed, the contract offered to the high-cost firm involves the price p_H. Assuming the rent of the high-cost firm, R_H, is zero, expression (4) implies the rent of the low-cost firm (again, conditional on the regulator being uninformed) is $\Delta^c Q(p_H)$.

The low-cost firm will find it too costly to bribe the informed regulator to conceal his information if

$$(1+\theta)s \geqslant \Delta^c Q(p_H), \tag{35}$$

where, recall, s is the payment from the principal to the regulator when the latter reports he has learned the firm has low cost. Expression (35) ensures the corruptible regulator is truthful when he announces he is ignorant about the firm's cost.

Suppose the regulator's income receives weight $\alpha_R \leqslant 1$ in the political principal's welfare function, while the rent of the firm has weight α. Then, much like expression (16), total expected welfare under this "collusion-proof" regulatory policy is

$$W = \psi\big[w_L(c_L) - (1-\alpha_R)s\big] + (1-\psi)\big[\phi^U\big\{w_L(c_L) - (1-\alpha)R_L\big\} + \big(1-\phi^U\big)w_H(p_H)\big].$$

Since payments from the political principal to the regulator are socially costly, inequality (35) will bind at the optimum. Consequently, total expected welfare is

$$W = \psi\left[w_L(c_L) - \frac{1-\alpha_R}{1+\theta}\Delta^c Q(p_H)\right] + (1-\psi)\big[\phi^U\big\{w_L(c_L) - (1-\alpha)\Delta^c Q(p_H)\big\} + \big(1-\phi^U\big)w_H(p_H)\big]. \tag{36}$$

Before deriving the price p_H that maximizes expected welfare, consider when the political principal will design the reward structure to ensure the regulator is not captured, i.e., when it is optimal to satisfy inequality (35). If the principal does not choose

[52] If explicit bribery were undertaken, θ might reflect the penalties associated with conviction for bribing an official, discounted by the probability of conviction.

s to satisfy (35), then the firm will always bribe the regulator to conceal damaging information, and so the regulator will never admit to being informed. In this case, the best the principal can do is follow the Baron–Myerson regulatory policy described in Proposition 1, where the policy designer has no additional private information. From expression (26), expected welfare in this case is

$$W = \phi\{w_L(c_L) - (1-\alpha)\Delta^c Q(p_H)\} + (1-\phi)w_H(p_H). \tag{37}$$

Using the identity $\psi + (1-\psi)\phi^U = \phi$, a comparison of welfare in (36) and (37) reveals that the political principal is better off using the corruptible (but sometimes well informed) regulator – and ensuring he is sufficiently well rewarded so as not to accept the firm's bribe – whenever $(1+\theta)(1-\alpha) > 1-\alpha_R$. In particular, if the regulator's rent receives at least as much weight in social welfare as the firm's rent, it is optimal to employ the regulator's information. Assume for the remainder of this section that this inequality holds.

Maximizing expression (36) with respect to p_H yields

$$p_H = \underbrace{c_H + \frac{\phi^U}{1-\phi^U}(1-\alpha)\Delta^c}_{\text{Baron–Myerson price}} + \underbrace{\frac{\psi}{(1-\psi)(1+\theta)(1-\phi^U)}(1-\alpha_R)\Delta^c}_{\text{extra distortion to reduce firm's stake in collusion}}. \tag{38}$$

From expression (5) in Proposition 1, when the regulator is uninformed and there is no scope for collusion, the optimal price for the high-cost firm is the first term in expression (38). The second term in (38) is an extra distortion in the high-cost firm's price that limits regulatory capture. The expression reveals that the danger of capture has no effect on optimal prices only when: (i) payments to the regulator have no social cost (i.e., when $\alpha_R = 1$), or (ii) it is prohibitively costly for the firm to bribe the regulator (so $\theta = \infty$).

The price for the high-cost firm is distorted further above cost when capture is possible because, from expression (35), a higher price for the high-cost firm reduces the rent that the low-cost firm would make if the informed regulator concealed his information. This reduced rent, in turn, reduces the bribe the firm will pay the regulator to conceal damaging information, which reduces the (socially costly) payment to the regulator that is needed to induce him to reveal his information. Most importantly, when there is a danger of regulatory capture, prices are distorted from their optimal levels when capture is not possible in a direction that reduces the firm's "stake in collusion", i.e., that reduces the rent the firm obtains when it captures the regulator. Interestingly, the possibility of capture – something that would clearly make the firm better off if the regulator were not adequately controlled – makes the firm worse off once the political principal has optimally responded to the threat of capture.

This discussion has considered what one might term the optimal response to the danger of capture and collusion.[53] We return to the general topic in Section 3.4.2, which focuses more on pragmatic responses to capture.

[53] In the same tradition, Laffont and Martimort (1999), building on Kofman and Lawarrée (1993), show how multiple regulators can act as a safeguard against capture when the "constitution" is designed optimally. In

2.4.3. Multi-dimensional private information

In practice, the regulated firm typically will have several pieces of private information, rather than the single piece of private information considered in the previous sections. For instance, a multiproduct firm may have private information about cost conditions for each of its products. Alternatively, a single-product firm may have privileged information about both its technology and consumer demand.

To analyze this situation formally, consider the following simple multiproduct extension of the Baron and Myerson (1982) model described in Section 2.3.1.[54] Suppose the firm supplies two products. The demand curve for each product is $Q(p)$ and demands for the two products are independent. The constant marginal cost of producing either product is either c_L or c_H.[55] The firm also incurs a known fixed cost, F. Thus, the firm can be one of four possible types, denoted $\{LL, LH, HL, HH\}$, where the type-ij firm incurs cost c_i in producing product 1 and cost c_j in producing product 2. Suppose the unconditional probability the firm has a low-cost realization for product 1 is ϕ. Let ϕ_i be the probability the firm has a low-cost realization for product 2, given that its cost is c_i for product 1. The cost realizations are positively correlated across products if $\phi_L > \phi_H$, negatively correlated if $\phi_L < \phi_H$, and statistically independent if $\phi_L = \phi_H$. To keep the analysis simple, suppose the unconditional probability of a low-cost realization for product 2 is also ϕ. In this case, states LH and HL are equally likely, so

$$\phi(1 - \phi_L) = (1 - \phi)\phi_H. \tag{39}$$

The regulator offers the firm a menu of options, so that if the firm announces its type to be ij, it must set the price p^1_{ij} for product 1, p^2_{ij} for product 2, and in return receive the transfer T_{ij}. The equilibrium rent of the type-ij firm under this policy is

$$R_{ij} = Q\big(p^1_{ij}\big)\big(p^1_{ij} - c_i\big) + Q\big(p^2_{ij}\big)\big(p^2_{ij} - c_j\big) - F + T_{ij}.$$

The participation constraints in the regulator's problem take the form $R_{ij} \geqslant 0$, of which only $R_{HH} \geqslant 0$ is relevant. (If the firm is one of the other three types, it can claim to have high cost for both products, and thereby make at least as much rent as R_{HH}.) There are twelve incentive compatibility constraints, since each of the four types of firm must have no incentive to claim to be any of the remaining three types. However, in this symmetric situation, one can restrict attention to only the constraints that ensure low-cost types do

the later paper, the presence of several regulators, each of whom observes a separate aspect of the firm's performance, relaxes relevant "collusion–proofness" constraints. The earlier paper focuses on the possibility that an honest regulator can observe when another regulator is corrupted, and so can act as a "whistle-blower".

[54] The following is based on Dana (1993) and Armstrong and Rochet (1999).

[55] Multi-dimensional private information is one area where the qualitative properties of the optimal regulatory policy can vary according to whether the firm's private information is discrete or continuous. One reason for the difference is that in a continuous framework it is generally optimal to terminate the operation of some firms in order to extract further rent from other firms. This feature can complicate the analysis (Armstrong, 1999, Section 2). Rochet and Stole (2003) survey the literature on multi-dimensional screening.

not claim to have high costs.[56] The symmetry of this problem ensures that only three rents are relevant: R_{HH}, R_{LL} and R_A. R_A is the firm's rent when its cost is high for one product and low for the other. ('A' stands for 'asymmetric'. We will refer to either the type-LH or the type-HL firm as the 'type-A' firm.) Similarly, there are only four prices that are relevant: p_{LL} is the price for both products when the firm has low cost for both products; p_{HH} is the price for both products when the firm has a high cost for both products; p_L^A is the price for the low-cost product when the firm has asymmetric costs, while p_H^A is the price for the high-cost product when the firm has asymmetric costs.

Much as in expression (16) for the single-product case, expected welfare in this setting is

$$\begin{aligned} W = {} & 2\phi(1-\phi_L)\left\{w_L\left(p_L^A\right) + w_H\left(p_H^A\right) - (1-\alpha)R_A\right\} \\ & + \phi\phi_L\left\{2w_L(p_{LL}) - (1-\alpha)R_{LL}\right\} \\ & + (1-\phi)(1-\phi_H)\left\{2w_H(p_{HH}) - (1-\alpha)R_{HH}\right\}. \end{aligned} \tag{40}$$

(Here, $w_i(p) = v(p) + Q(p)(p - c_i)$, where $v(\cdot)$ again denotes consumer surplus.) The incentive compatibility constraint that ensures the type-A firm does not claim to be the type-HH firm is

$$\begin{aligned} R_A & \geqslant Q(p_{HH})(p_{HH} - c_H) + Q(p_{HH})(p_{HH} - c_L) - F + T_{HH} \\ & = R_{HH} + \Delta^c Q(p_{HH}), \end{aligned} \tag{41}$$

where $\Delta^c = c_H - c_L$. Similarly, the incentive compatibility constraint that ensures the type-LL firm does not claim to be a type-A firm is

$$R_{LL} \geqslant R_A + \Delta^c Q\left(p_H^A\right). \tag{42}$$

Finally, the incentive compatibility constraint that ensures the type-LL firm does not claim to be a type-HH firm is

$$R_{LL} \geqslant R_{HH} + 2\Delta^c Q(p_{HH}). \tag{43}$$

The participation constraint for the type-HH firm will bind, so $R_{HH} = 0$. The type-A firm's incentive compatibility constraint (41) will also bind, so $R_A = \Delta^c Q(p_{HH})$. Substituting these rents into (42) and (43) implies the rent of the type-LL firm is

$$R_{LL} = \Delta^c Q(p_{HH}) + \max\left\{\Delta^c Q\left(p_H^A\right), \Delta^c Q(p_{HH})\right\}.$$

Substituting these rents into expected welfare (40) implies welfare is

$$W = 2\phi(1-\phi_L)\left\{w_L\left(p_L^A\right) + w_H\left(p_H^A\right) - (1-\alpha)\Delta^c Q(p_{HH})\right\}$$

[56] It is straightforward to verify that the other incentive compatibility constraints are satisfied at the solution to the regulator's problem in this symmetric setting. Armstrong and Rochet (1999) show that in the presence of negative correlation and substantial asymmetry across markets, some of the other incentive compatibility constraints may bind at the solution to the regulator's problem, and so cannot be ignored in solving the problem.

$$+ \phi\phi_L\{2w_L(p_{LL}) - (1-\alpha)2\Delta^c Q(p_{HH})\}$$
$$+ (1-\phi)(1-\phi_H)2w_H(p_{HH}) \tag{44}$$

if $p_H^A \geqslant p_{HH}$, and

$$\begin{aligned} W = {} & 2\phi(1-\phi_L)\{w_L(p_L^A) + w_H(p_H^A) - (1-\alpha)\Delta^c Q(p_{HH})\} \\ & + \phi\phi_L\{2w_L(p_{LL}) - (1-\alpha)\Delta^c[Q(p_{HH}) \\ & + Q(p_H^A)]\} + (1-\phi)(1-\phi_H)2w_H(p_{HH}) \end{aligned} \tag{45}$$

if $p_H^A \leqslant p_{HH}$.

The policy that maximizes welfare consists of the prices $\{p_{LL}, p_{HH}, p_L^A, p_H^A\}$ that maximize the expression in (44)–(45). Some features of the optimal policy are immediate. First, since the prices for low-cost products (p_{LL} and p_L^A) do not affect any rents, they are not distorted, and are set equal to marginal cost c_L. This generalizes Proposition 1.[57] Second, the strict inequality $p_H^A > p_{HH}$ cannot be optimal. To see why, notice that when this inequality holds, expression (44) is the relevant expression for welfare. In expression (44), price p_H^A does not affect rent. Consequently, $p_H^A = c_H$ is optimal. But the value of p_{HH} that maximizes expression (44) exceeds cost c_H. Therefore, the inequality $p_H^A \geqslant p_{HH}$ must bind if (44) is maximized subject to the constraint $p_H^A \geqslant p_{HH}$. In sum, attention can be restricted to the case where $p_H^A \leqslant p_{HH}$, and so (45) is the appropriate expression for welfare.

The remaining question is whether $p_H^A = p_{HH}$ or $p_H^A < p_{HH}$ is optimal. If the constraint that $p_H^A \leqslant p_{HH}$ is ignored, the prices that maximize (45) are:

$$p_H^A \quad \text{maximizes} \quad 2w_H(\cdot) - \frac{\phi_L}{1-\phi_L}(1-\alpha)\Delta^c Q(\cdot), \quad \text{and} \tag{46}$$

$$p_{HH} \quad \text{maximizes} \quad 2w_H(\cdot) - \frac{1-(1-\phi)(1-\phi_H)}{(1-\phi)(1-\phi_H)}(1-\alpha)\Delta^c Q(\cdot). \tag{47}$$

Clearly, the price p_{HH} in (47) exceeds the price p_H^A in (46) whenever

$$\phi_L \leqslant 1-(1-\phi)(1-\phi_H), \tag{48}$$

which is equivalent to the condition $\phi_L \leqslant 2\phi_H$.[58] This inequality states that the correlation between the cost realizations is not too pronounced. When this condition is satisfied, expressions (46) and (47) provide the two high-cost prices.

[57] Armstrong and Rochet (1999) show that when there is negative correlation and conditions are very asymmetric across the two markets, it is optimal to introduce distortions even for efficient firms. The distortions take the form of below-cost prices.

[58] This non-trivial manipulation involves using expression (39) to write $\phi = \phi_H/(1+\phi_H-\phi_L)$, and substituting this into inequality (48).

When there is strong positive correlation, so $\phi_L \geqslant 2\phi_H$, the constraint $p_H^A \leqslant p_{HH}$ binds. Letting $p_H = p_H^A = p_{HH}$ denote this common price, (45) simplifies to

$$W = 2\{\phi[w_L(c_L) - (1-\alpha)\Delta^c Q(p_H)] + (1-\phi)w_H(p_H)\},$$

which is just (twice) the standard single-product Baron–Myerson formula. (See expression (26) for instance.) Therefore, with strong positive correlation, the solution is simply the Baron–Myerson formula (5) for each product. This discussion constitutes the proof of Proposition 5.[59]

PROPOSITION 5. *The optimal policy in the symmetric multi-dimensional setting has the following features*:

(i) *There are no pricing distortions for low-cost products, i.e.,* $p_{LL} = p_L^A = c_L$.

(ii) *When there is strong positive correlation between costs (so* $\phi_L \geqslant 2\phi_H$*), the regulatory policy for each product is independent of the firm's report for the other product. The policy for each product is identical to the policy described in Proposition 1.*

(iii) *When cost correlation is weak (so* $\phi_L \leqslant 2\phi_H$*), interdependencies are introduced across products. In particular,*

$$p_H^A = c_H + \frac{\phi_L}{2(1-\phi_L)}(1-\alpha)\Delta^c, \quad \textit{and}$$
$$p_{HH} = c_H + \frac{1-(1-\phi)(1-\phi_H)}{2(1-\phi)(1-\phi_H)}(1-\alpha)\Delta^c \geqslant p_H^A.$$

Part (i) of Proposition 5 provides the standard conclusion that price is set equal to cost whenever the low cost is realized. Since the binding constraint is to prevent the firm from exaggerating, not understating, its costs, no purpose would be served by distorting prices when low costs are reported. Part (ii) provides another finding that parallels standard conclusions. It states that in the presence of strong positive cost correlation, the optimal policy is the same for each product and depends only on the realized cost of producing that product. Furthermore, this optimal policy replicates the policy that is implemented in the case of uni-dimensional cost uncertainty, as described in Proposition 1. Thus, in the presence of strong correlation, the two-dimensional problem essentially is transformed into two separate uni-dimensional problems. The reason for this result is the following. When there is strong positive correlation, the most likely realizations are type *LL* and type *HH*. Consequently, the most important incentive compatibility constraint is that the type-*LL* firm should not claim to be type *HH*. This problem is analogous to the single-product Baron–Myerson problem, and so the optimal policy in this two-dimensional setting parallels the optimal policy in the uni-dimensional Baron–Myerson setting.

[59] In the knife-edge case where $\phi_L = 2\phi_H$, the policies in parts (ii) and (iii) of this proposition generate the same optimal welfare for the regulator.

Part (iii) reveals a major difference between the two-dimensional and uni-dimensional settings. It states that in the presence of weak cost correlation, when the firm has a high cost for one product, its price is set closer to cost for that product when its cost is low for the other product than when its cost is high for the other product. The less pronounced distortion when the asymmetric pair of costs $\{c_L, c_H\}$ is realized is optimal because this realization is relatively likely with weak cost correlation.[60] In contrast, the simultaneous realization of high cost for both products is relatively unlikely. So the expected loss in welfare from setting p_{HH} well above cost c_H is small. Furthermore, this distortion reduces the attraction to the firm of claiming to have high cost for both products in the relatively likely event that the firm has high cost for one product and low cost for the other.

A second regulatory setting in which the firm's superior information is likely to be multi-dimensional occurs when the firm is privately informed about both its cost structure and the consumer demand for its product. Private cost and demand information enter the analysis in fundamentally asymmetric ways. Consequently, this analysis is more complex than the analysis reviewed above. It can be shown that it is sometimes optimal to require the regulated firm to set a price below its realized cost when the firm is privately informed about both its demand and cost functions. Setting a price below marginal cost can help discourage the firm from exaggerating the scale of consumer demand.[61]

2.5. Dynamic interactions

Now consider how optimal regulatory policy changes when the interaction between the regulator and the regulated firm is repeated. To do so most simply, suppose their interaction is repeated just once in the setting where the firm is privately informed about its unobservable, exogenous marginal cost of production. We will employ notation similar to that used in Section 2.3.1. For simplicity, suppose the firm's cost $c \in \{c_L, c_H\}$ is perfectly correlated across the two periods.[62] Let $\phi \in (0, 1)$ be the probability the firm has low marginal cost, c_L, in the two periods. The regulator and the firm have the same discount factor $\delta > 0$. The demand function in the two periods, $Q(p)$, is common knowledge. The regulator wishes to maximize the expected discounted weighted sum of consumer surplus and rent. The firm will only produce in the second period if it receives non-negative rent from doing so, just as it will only produce in the first period if it anticipates non-negative expected discounted rent from doing so.

[60] Notice from part (iii) that in the extreme case where the type-LL realization never occurs, i.e., when $\phi_L = 0$, the prices of the type-A firm will not be distorted.

[61] See Lewis and Sappington (1988b) and Armstrong (1999) for analyses of this problem.

[62] See Bolton and Dewatripont (2005, Section 9.1.4) for a similar account of regulatory dynamics in the context of the Laffont and Tirole (1993b) model. See Baron and Besanko (1984a) and Laffont and Martimort (2002, Section 8.1.3) for an analysis of the case where the firm's costs are imperfectly correlated over time and where the regulator's commitment powers are unimpeded.

The ensuing discussion analyzes three variants of dynamic regulation that differ according to the commitment abilities of the regulator. The discussion is arranged in order of decreasing commitment power for the regulator.

2.5.1. Perfect intertemporal commitment

This first case is the most favorable one for the regulator because he can commit to any dynamic regulatory policy. In this case, the regulator will offer the firm a long-term (two-period) contract. The regulatory policy consists of a pair of price and transfer payment options $\{(p_L, T_L), (p_H, T_H)\}$ from which the firm can choose. In principle, these options could differ in the two time periods. However, it is readily verified that such variation is not optimal when costs do not vary over time. Consequently, the analysis in this two-period setting with perfect intertemporal regulatory commitment parallels the static analysis of Proposition 1, and the optimal dynamic policy simply duplicates the single-period policy in each period.

PROPOSITION 6. *In the two-period setting with regulatory commitment, the optimal regulatory policy has the following features*:

(i) *Prices in each of the two periods are*[63]

$$p_L = c_L, \qquad p_H = c_H + \frac{\phi}{1-\phi}(1-\alpha)\Delta^c.$$

(ii) *Total discounted rents are*

$$R_L = (1+\delta)\Delta^c Q(P_H), \qquad R_H = 0.$$

Once the regulator has observed the choice made by the firm in the first period, he would like to change the second-period policy to rectify the two undesirable features of the optimal regulatory policy under asymmetric information. Recall from Proposition 1 that the high-cost firm charges a price that is distorted above its marginal cost and the low-cost firm obtains a socially costly rent. By the second period, the regulator is fully informed about the firm's cost. Therefore, if the firm has revealed it has high cost, the regulator would like to reduce the firm's price to the level of its cost. Here, the temptation is not to eliminate rent (i.e., to "expropriate" the firm), but rather to achieve more efficient pricing. In this case, therefore, there is scope for mutually beneficial modifications to the pre-specified policy. Alternatively, if the firm has revealed it has low cost, the regulator would like to keep the price the same but eliminate the firm's rent. In this instance, the danger is that the regulator may expropriate the firm. Such a change in regulatory policy would not be mutually improving. These two temptations are the subject of the two commitment problems discussed next.

[63] This is a special case of part (ii) of Proposition 5.

The regulator with full commitment power does not succumb to these temptations. Instead, the regulator optimally commits not to use against the firm in the second period any cost information he infers from the firm's first-period actions. The regulator does this in order to best limit the rent that accrues to the firm with low cost.

2.5.2. Long-term contracts: the danger of renegotiation

Now consider the case where the regulator has "moderate" commitment powers.[64] Specifically, the regulator and the firm can write binding long-term contracts, but they cannot commit not to renegotiate the original contract if both parties agree to do so (i.e., if there is scope for Pareto gains ex post). Thus, the regulator cannot credibly promise to leave in place a policy that he believes to be Pareto inefficient in the light of information revealed to him. However, the regulator can credibly promise not to use information he has obtained to eliminate the firm's rent. In particular, because a policy change requires the consent of both parties, the regulator cannot reduce the rent of the low-cost firm below the level of rent it would secure if it continued to operate under the policy initially announced by the regulator.

In essence, this renegotiation setting presumes that the regulator can commit to provide specified future rent to the firm, but not to how that rent will be delivered (i.e., to the particular prices and transfers that generate the rent). The firm is indifferent as to how its rent is generated. However, the composition of rent affects the firm's incentives to reveal its cost truthfully.

The optimal policy with full commitment (Proposition 6) is no longer possible with renegotiation. The fact that the firm chose p_H initially implies that it has high cost in the second period, and, therefore, that mutual gains could be secured by reducing price to c_H in the second period. In the renegotiation setting, then, whenever definitive cost information is revealed in the first period, the regulator will always charge marginal-cost prices in the second period. It is apparent that this policy is not ideal for the regulator, since the regulator with full commitment powers could implement this policy, but chooses not to do so.

Formally, activity in the renegotiation setting proceeds as follows. First, the regulator announces the policy that will be implemented in the first period and the policy that, unless altered by mutual consent, will be implemented in the second period. Second, the firm chooses its preferred first-period option from the options presented by the regulator. After observing the firm's first-period activities and updating his beliefs about the firm's capabilities accordingly, the regulator can propose a change to the policy he announced initially.[65] If he proposes a change, the firm then decides whether to accept the change.

[64] This discussion is based on Laffont and Tirole (1990a) and Laffont and Tirole (1993b, ch. 10). For an alternative model of moderate commitment power, see Baron and Besanko (1987a).

[65] All parties can anticipate fully any modification of the original policy that the regulator will ultimately propose. Consequently, there is no loss of generality in restricting attention to renegotiation-proof policies, which are policies to which the regulator will propose no changes once they are implemented. See Laffont and Tirole (1993b, pp. 443–447) for further discussion of this issue.

If the firm agrees to the change, it is implemented. If the firm does not accept the change, the terms of the original policy remain in effect.

It is useful as a preliminary step to derive the optimal *separating* contracts in the renegotiation setting, that is to say, the optimal contracts that fully reveal the firm's private information in the first period. Suppose the regulator offers the type-i firm a long-term contact such that, in period 1 the firm charges the price p_i and receives the transfer T_i, and in the second period the firm is promised a rent equal to R_i^2. In this case, given discount factor δ, the total discounted rent of the type-i firm is

$$R_i = Q(p_i)(p_i - c_i) - F + T_i + \delta R_i^2.$$

By assumption, the firm's cost level is fully revealed by its choice of first-period contract. Because the regulator will always provide the promised second-period rent in the most efficient manner, he will set the type i firm's second-period price equal to c_i and implement the transfer payment that delivers rent R_i^2. Therefore, the incentive compatibility constraint for the low-cost firm, when it foresees that the second-period price will be c_H if it claims to have high cost, is

$$\begin{aligned} R_L &\geqslant Q(p_H)(p_H - c_L) - F + T_H + \delta\big\{Q(c_H)(c_H - c_L) + R_H^2\big\} \\ &= R_H + \big[Q(p_H) + \delta Q(c_H)\big]\Delta^c. \end{aligned} \tag{49}$$

If the incentive compatibility constraint (49) binds and the participation constraint of the high-cost firm binds (so $R_H = 0$), then total discounted welfare is

$$\begin{aligned} W = {}& \phi\big\{w_L(p_L) + \delta w_L(c_L) - (1-\alpha)\Delta^c\big[Q(p_H) + \delta Q(c_H)\big]\big\} \\ &+ (1-\phi)\big\{w_H(p_H) + \delta w_H(c_H)\big\}. \end{aligned} \tag{50}$$

Maximizing expression (50) with respect to the remaining choice variables, p_L and p_H, reveals that the first-period prices are precisely those identified in Proposition 1 (and hence also those in part (i) of Proposition 6). Notice in particular that when separation is induced, first-period prices are not affected by the regulator's limited commitment powers. Limited commitment simply forces the regulator to give the low-cost firm more rent.

It is useful to decompose the expression for welfare in (50) into the welfare achieved in the first period and the welfare achieved in the second period. Doing so reveals

$$\begin{aligned} W = {}& \underbrace{\phi\big\{w_L(c_L) - (1-\alpha)\Delta^c Q(p_H)\big\} + (1-\phi)w_H(p_H)}_{\text{welfare from Baron–Myerson regime}} \\ &+ \delta \underbrace{\big[\phi\big\{w_L(c_L) - (1-\alpha)\Delta^c Q(c_H)\big\} + (1-\phi)w_H(c_H)\big]}_{\text{welfare from Loeb–Magat regime}}. \end{aligned} \tag{51}$$

Since price p_H in expression (51) is the optimal static price in Proposition 1, welfare in the first period is precisely that achieved by the Baron–Myerson solution to the static problem. Because both prices are set equal to cost in the second period when separation is induced, second-period welfare is the welfare achieved when both firms offer marginal-cost prices, and the low-cost firm is offered the high rent ($\Delta^c Q(c_H)$) to ensure incentive compatibility.[66] This second-period policy is not optimal, except in the extreme setting where $\alpha = 1$, in which case intertemporal commitment power brings no benefit for regulation. The reduced welfare represents the cost that arises (when separation is optimal) from the regulator's inability to commit not to renegotiate.

However, the optimal regulatory policy will not always involve complete separation in the first period.[67] To see why most simply, consider the discounted welfare resulting from a policy of complete pooling in the first period. Under the optimal pooling contract, the firm charges the same price $\tilde{p}$, say, in the first period, regardless of its cost. The high-cost firm obtains zero rent and the low-cost firm obtains rent $\Delta^c Q(\tilde{p})$ in the first period. Clearly, such a policy yields lower welfare than the level derived in the Baron–Myerson regime in the first period. However, it has the benefit that at the start of the second period the regulator has learned nothing about the firm's realized cost, and so there is no scope for renegotiation. In particular, in the second period, the optimal policy will be precisely the Baron–Myerson policy of Proposition 1.

Thus, compared to the optimal separating equilibrium in (51), the pooling regime results in lower welfare in the first period and higher welfare in the second. Much as in expression (51), total discounted welfare under this policy is

$$W = \underbrace{\phi\big\{w_L(\tilde{p}) - (1-\alpha)\Delta^c Q(\tilde{p})\big\} + (1-\phi)w_H(\tilde{p})}_{\text{welfare from pooling regime}}$$
$$+ \delta \underbrace{\big[\phi\big\{w_L(c_L) - (1-\alpha)\Delta^c Q(p_H)\big\} + (1-\phi)w_H(p_H)\big]}_{\text{welfare from Baron–Myerson regime}}.$$

Whenever the discount factor δ is sufficiently large (i.e., substantially greater than unity), the second-period welfare gains resulting from first-period pooling will outweigh the corresponding first-period losses, and a separating regulatory policy is not optimal. A pooling policy in the first period can be viewed as a (costly) means by which the regulator can increase his commitment power.

Thus, some pooling will optimally be implemented whenever the regulator and the firm value the future sufficiently highly.[68] When separation is not optimal, the precise details of the optimum are intricate. In rough terms, when the discount factor δ is small

[66] Recall the discussion in Section 2.3.1 of the policy suggested by Loeb and Magat (1979).

[67] In fact, when private information is distributed continuously (not discretely, as presumed in this chapter), a fully-separating first-period set of contracts is never optimal (although it is feasible) – see Laffont and Tirole (1993b, Section 10.6).

[68] Complete pooling is never optimal for the regulator. Reducing the probability that the two types of firm are pooled to slightly below 1 provides a first-order gain in first-period welfare by expanding the output of the low-cost firm toward its efficient level. Any corresponding reduction in second-period welfare is of second-order magnitude because, with complete pooling, the optimal second-period regulatory policy is precisely the policy that is optimal in the single-period setting when ϕ is the probability that the firm has low costs.

enough, the separation contracts derived above are optimal. As δ increases, a degree of pooling is optimal and the amount of pooling increases with δ.[69]

This particular commitment problem is potentially hard to overcome because it arises simply from the possibility that the regulator and firm mutually agree to alter the terms of a prevailing contract. In practice, an additional problem is that political pressure from consumer advocates, for example, might make it difficult for the regulator knowingly to continue to deliver rent to the regulated firm. This problem is discussed next.

2.5.3. *Short-term contracts: the danger of expropriation*

Now consider the two-period setting of Section 2.5.2 with one exception: the regulator cannot credibly commit to deliver specified second-period rents.[70] In other words, the regulator cannot specify the policy he will implement in the second period until the start of that period. In this case, the low-cost firm will be reluctant to reveal its superior capabilities, since such revelation will eliminate its second-period rent. In contrast to the renegotiation model, there are no long-term contracts in this setting that can protect the firm against such expropriation.

The optimal separating regulatory policy in the no-commitment setting can be derived much as it was derived in the renegotiation setting of Section 2.5.2. Suppose the regulator offers the two distinct options (p_L, T_L) and (p_H, T_H) in the first period, and the type-i firm chooses the (p_i, T_i) option with probability one. Because the firm's first-period choice fully reveals its second-period cost, second-period prices will be set equal to marginal costs, and the transfer will be set equal to the firm's fixed cost of production. Because the firm never receives any rent in the second period in this separating equilibrium, the rent of the type-i firm over the two periods is $R_i = Q(p_i)(p_i - c_i) - F + T_i$. Therefore, to prevent the low-cost firm from exaggerating its cost in the first period, it must be the case that

$$\begin{aligned} R_L &\geqslant Q(p_H)(p_H - c_L) - F + T_H + \delta \Delta^c Q(c_H) \\ &= R_H + \left[Q(p_H) + \delta Q(c_H)\right]\Delta^c. \end{aligned} \tag{52}$$

Thus, the low-cost firm must be promised a relatively large first-period rent, R_L, to induce it to reveal its superior capabilities. Notice that expression (52) is precisely the

[69] See Laffont and Tirole (1993b, ch. 10) for details of the solution. Notice that the revelation principle is no longer valid in dynamic settings without commitment. That is to say, the regulator may do better if he considers contracts other than those where the firm always reveals its type. See Bester and Strausz (2001) for a precise characterization of optimal contracts without commitment. (Laffont and Tirole did not consider all possible contracts (see p. 390 of their book), but Bester and Strausz show that the contracts Laffont and Tirole consider include the optimal contracts.) For additional analysis of the design of contracts in the presence of adverse selection and renegotiation, see Rey and Salanié (1996), for example.

[70] This discussion is based on Laffont and Tirole (1988a) and Laffont and Tirole (1993b, ch. 9). Freixas, Guesnerie and Tirole (1985) explore a related model that considers linear contracts.

incentive compatibility constraint (49) for the low-cost firm in the setting with renegotiation. Assuming incentive constraint (52) binds and the participation constraint for the high-cost firm binds, welfare is given by expressions (50) and (51). Natural candidates for optimal first-period prices are derived by maximizing this expression with respect to p_L and p_H, which provides the prices identified in Proposition 1.

However, in contrast to the static analysis (and the renegotiation analysis), it is not always appropriate to ignore the high-cost firm's incentive compatibility constraint when the regulator has no intertemporal commitment powers. This constraint may be violated if the firm can refuse to produce in the second period without penalty. In this case, the high-cost firm may find it profitable to understate its first-period cost, collect the large transfer payment intended for the low-cost firm, and then terminate second-period operations rather than sell output in the second period at a price (c_L) below its cost c_H.[71]

To determine when the incentive compatibility constraint for the high-cost firm binds, notice that when the constraint is ignored and $R_H = 0$, the regulator optimally sets $p_L = c_L$ and $T_L = [Q(p_H) + \delta Q(c_H)]\Delta^c + F$. Consequently, the high-cost firm will not find it profitable to understate its cost under this regulatory policy if

$$0 \geqslant Q(c_L)(c_L - c_H) - F + T_L = \big[Q(p_H) + \delta Q(c_H) - Q(c_L)\big]\Delta^c. \tag{53}$$

When p_H is as specified in Equation (5) in Proposition 1, expression (53) will hold as a strict inequality when the discount factor δ is sufficiently small. Therefore, for small δ, the identified regulatory policy is the optimal one when the regulator cannot credibly commit to future policy.[72] Just as in the renegotiation setting, first-period prices are not affected by the regulator's limited commitment powers. Limited commitment simply forces the regulator to compensate the low-cost firm in advance for the second-period rent it foregoes by revealing its superior capabilities in the first period.

When the regulator and firm do not discount the future highly, inequality (53) will not hold, and so the incentive compatibility constraint for the high-cost firm may bind. To relax this constraint, the regulator optimally increases the incremental first-period output $(Q(p_L) - Q(p_H))$ the firm must deliver when it claims to have low cost. This increase is accomplished by reducing p_L below c_L and raising p_H above the level identified in expression (5) of Proposition 1. The increased output when low cost is reported reduces the profit of the high-cost firm when it understates its cost. The profit reduction arises

[71] Laffont and Tirole call this the "take the money and run" strategy. This possibility is one of the chief differences between the setting with renegotiation and the setting with no commitment. Under renegotiation, transfers and rents can be structured over time so that this strategy is never profitable for the high-cost firm. In particular, the renegotiation model gives rise to a more standard structure (i.e., the "usual" incentive compatibility constraints bind) than the no-commitment model.

[72] When private information is distributed continuously (rather than discretely as presumed in this chapter), it is never feasible (let alone optimal) to have a fully-revealing first-period set of contracts. Because the firm obtains zero rent in the second period under any contract that induces full separation in the first period, a firm would always find it profitable to mimic a slightly less efficient firm. This deviation will introduce only a second-order reduction in rent in the first period, but a first-order increase in rent in the second period. See Laffont and Tirole (1993b, Section 9.3).

because the corresponding increase in the transfer payment is only c_L per unit of output, which is compensatory for the low-cost firm, but not for the high-cost firm.

Although these distortions limit the firm's incentive to understate its cost, they also reduce total surplus. Beyond some point, the surplus reduction resulting from the distortions required to prevent cost misrepresentation outweigh the potential gains from matching the second-period price to the realized marginal cost. Consequently, the regulator will no longer ensure the low-cost and high-cost firm always set distinct prices. Instead, the regulator will induce the distinct types of the firm to implement the same price in the first period with positive probability (i.e., partial pooling is implemented).

These conclusions are summarized in Proposition 7.

PROPOSITION 7. *In the two-period setting with no regulatory commitment, the optimal regulatory policy has the following features*:

(i) *When δ is sufficiently small that inequality (53) holds, the prices identified in Proposition 1 are implemented in the first period, and the full-information outcome is implemented in the second period.*

(ii) *For larger values of δ, if separation is induced in the first period, p_L is set below c_L and p_H is set above the level identified in Proposition 1. The full-information outcome is implemented in the second period.*

(iii) *When δ is sufficiently large, partial pooling is induced in the first period.*

The pooling identified in property (iii) of Proposition 7 illustrates an important principle.[73] When regulators cannot make binding commitments regarding their use of pertinent information, welfare may be higher when regulators are denied access to the information. To illustrate, when a regulator cannot refrain from matching prices to realized production costs, welfare can increase as the regulator's ability to monitor realized production costs declines. When the regulator is unable to detect realized cost reductions immediately, the firm's incentives to deliver the effort required in order to reduce cost are enhanced. As a result, profit and consumer surplus can both increase.[74]

Another important feature of the outcome with no commitment (and also with renegotiation) is that, at least when δ is sufficiently small that first-period separation is optimal, the firm benefits from the regulator's limited commitment powers. One might expect that a regulator's inability to prevent himself from expropriating the firm's rents would make the firm worse off. However, notice that the high-cost firm makes no rent whether

[73] Notice that a lack of intertemporal commitment presents no problems for regulation when the static problem involves complete pooling (as is the case, for instance, when demand is unknown and the firm has a concave cost function). At the other extreme, when the full-information optimum is feasible in the static problem (e.g., when demand is unknown and the cost function is convex) there is no further scope for expropriation in the second period. Consequently, the regulator again does not need any commitment ability to achieve the ideal outcome in this dynamic context.

[74] See Sappington (1986). This conclusion is closely related to the principles that inform the optimal length of time between regulatory reviews of the firm's performance. See Section 3.2.3.

the regulator's commitment powers are limited or unlimited, and so is indifferent between the two regimes. Without commitment, expression (52) reveals that the low-cost firm makes discounted rent $[Q(p_H) + \delta Q(c_H)]\Delta^c$. With commitment, Proposition 6 reveals that the corresponding rent is only $[Q(p_H) + \delta Q(p_H)]\Delta^c$. Because $p_H > c_H$ and so $Q(c_H) > Q(p_H)$, the firm gains when the regulator cannot credibly promise to refrain from expropriating the firm.

Of course, in practice a regulator can exploit the firm's sunk physical investments as well as information about the firm's capabilities. We return to the general topic of policy credibility and regulatory expropriation in Section 3.4.1.

2.6. *Regulation under moral hazard*

To this point, the analysis has focused on the case where the firm is perfectly informed from the outset about its exogenous production cost. In practice, a regulated firm often will be uncertain about the operating cost it can achieve, but knows that it can reduce expected operating cost by undertaking cost-reducing effort. The analysis in this section considers how the regulator can best motivate the firm to deliver such unobservable cost-reducing effort.[75]

The simple moral hazard setting considered here parallels the framework of Section 2.3.3 where there are two states, denoted L and H (which could denote different technologies or different demands, for example). State L is the socially desirable state. As before, let $\phi \in (0, 1)$ be the probability that state L is realized. However, the parameter ϕ is chosen by the firm in the present setting. The increasing, strictly convex function $D(\phi) \geqslant 0$ denotes the disutility incurred by the firm in securing the probability ϕ. The regulator cannot observe the firm's choice of ϕ, which can be thought of as the firm's effort in securing the favorable L state. The regulator can accurately observe the realized state, and offers the firm a pair of utilities, $\{U_L, U_H\}$, where the firm enjoys the utility U_i when state i is realized.[76] Because of the uncertainty of the outcome, the firm's attitude towards risk is important, and so we distinguish between 'utility' and 'rent'. (In the special case where the firm is risk neutral, the two concepts coincide.)

The firm's expected utility when it delivers the effort required to ensure success probability ϕ (i.e., to ensure that state L occurs with probability ϕ) is

$$U = \phi U_L + (1 - \phi) U_H - D(\phi) \geqslant U^0, \tag{54}$$

where the inequality in expression (54) indicates that the firm must achieve expected utility of at least U^0 if it is to be willing to participate. The firm's optimal choice of

[75] We are unaware of a treatment of the regulatory moral hazard problem that parallels exactly the problem analyzed in this section. For recent related discussions of the moral hazard problem, see Laffont and Martimort (2002, chs. 4 and 5) and Bolton and Dewatripont (2005, ch. 4). See Cowan (2004) for an analysis of optimal risk-sharing between consumers and the regulated firm in a full information framework.

[76] If the regulator could not observe the realized state in this setting, an adverse selection problem would accompany the moral hazard problem. See Laffont and Martimort (2002, Section 7.2) for an analysis of such models.

ϕ can be expressed as a function of the incremental utility it anticipates in state L, $\Delta^U = U_L - U_H$. The magnitude of Δ^U represents the *power* of the incentive scheme used to motivate the firm. Formally, from expression (54) the firm's equilibrium level of effort, denoted $\hat{\phi}(\Delta^U)$, satisfies:

$$D'\big(\hat{\phi}\big(\Delta^U\big)\big) \equiv \Delta^U. \tag{55}$$

Equilibrium effort $\hat{\phi}$ is an increasing function of the power of the incentive scheme, Δ^U.

For simplicity, suppose the regulator seeks to maximize expected consumer surplus.[77] Suppose that in state i, if the firm is given utility U_i, the maximum level of consumer surplus available is $V_i(U_i)$. (We will illustrate this relationship between consumer surplus and the firm's utility shortly.) Therefore, the regulator wishes to maximize

$$V = \phi V_L(U_L) + (1 - \phi)V_H(U_H),$$

subject to the participation constraint (54) and the equilibrium effort condition $\phi = \hat{\phi}(\Delta^U)$ as defined by expression (55). Given the presumed separability in the firm's utility function, the participation constraint (54) will bind at the optimum. Therefore, we can re-state the regulator's problem as maximizing social surplus

$$W = \phi W_L(U_L) + (1 - \phi)W_H(U_H) - D(\phi), \tag{56}$$

where $W_i(U_i) \equiv V_i(U_i) + U_i$, subject to $\phi = \hat{\phi}(\Delta^U)$ and the participation constraint (54).

We next describe three natural examples of the relationship $V_i(U_i)$ between the firm's utility and consumer surplus. In each of these examples, suppose the firm's profit in state i is $\pi_i(p_i)$ when it offers the price p_i, and $v_i(p_i)$ is (gross) consumer surplus. Let $w_i(\cdot) \equiv v_i(\cdot) + \pi_i(\cdot)$ denote the total unweighted surplus function, and suppose p_i^* is the price that maximizes welfare $w_i(\cdot)$ in state i. If the regulator requires the firm to offer the price p_i and gives the firm a transfer payment T_i in state i, the rent of the firm is $R_i = \pi_i(p_i) + T_i$.[78]

Case 1: Risk-neutral firm when transfers are employed

When the firm is risk neutral its utility is equal to its rent, and so $U_i = R_i = \pi_i(p_i) + T_i$. Therefore, $V_i(U_i)$, which is the maximum level of (net) consumer surplus $v_i(p_i) - T_i$ that can be achieved for a utility level U_i, is given by

$$V_i(U_i) = w_i\big(p_i^*\big) - U_i.$$

[77] Thus, we assume consumers are "risk neutral" in their valuation of consumer surplus. The ensuing analysis is unaltered if the regulator seeks to maximize a weighted average of consumer surplus and utility, $S + \alpha U$, provided the weight α is not so large that the firm's participation constraint does not bind at the optimum.

[78] For ease of exposition, we assume the firm produces a single product. The analysis is readily extended to allow for multiple products.

In this case, the firm's utility and maximized consumer surplus sum to a constant, i.e.,

$$W_i(U_i) \equiv w_i\big(p_i^*\big), \tag{57}$$

and the available total surplus is invariant to the rent/utility afforded the firm.

Case 2: Risk-averse firm when transfers are employed

When the firm is risk averse and its rent in state i is $R_i = \pi_i(p_i) + T_i$, its utility U_i can be written as $u(R_i)$ where $u(\cdot)$ is a concave function. Therefore, $V_i(U_i)$ is given by

$$V_i(U_i) = w_i\big(p_i^*\big) - u^{-1}(U_i), \tag{58}$$

where $u^{-1}(\cdot)$ is the inverse function of $u(\cdot)$. Here, there is a decreasing and concave relationship between firm utility and consumer surplus. In this case, firm utility and maximized consumer surplus do not sum to a constant, and $W_i(U_i)$ is a concave function. However, the trade-off between the firm's utility and consumer surplus does not depend on the prevailing state. Consequently,

$$V_L'(U) \equiv V_H'(U). \tag{59}$$

Case 3: Risk-neutral firm when no transfers are employed

When the firm is risk neutral and no lump-sum transfers are employed, $U_i = R_i = \pi_i(p_i)$. Therefore, $V_i(U_i)$ is just the level of consumer surplus $v_i(p_i)$ when the price is such that $\pi_i(p_i) = U_i$. Consequently,

$$V_i(U_i) = v_i\big(\pi_i^{-1}(U_i)\big). \tag{60}$$

In this case, firm utility and maximized consumer surplus again do not sum to a constant. In the special case where the demand function is iso-elastic with elasticity η,

$$V_i'(U_i) = \frac{-1}{1 - \eta[\frac{p_i - c_i}{p_i}]}, \tag{61}$$

where p_i is the price that yields rent $U_i = \pi_i(p_i)$.

Full-information benchmark Consider the hypothetical setting where the regulator can directly control effort ϕ, so the effort selection constraint, $\phi = \hat{\phi}(\Delta^U)$, can be ignored. If λ is the Lagrange multiplier for the participation constraint (54) in this full-information problem, the optimal choices for U_L and U_H satisfy

$$V_L'(U_L) = V_H'(U_H) = -(1 + \lambda). \tag{62}$$

Expression (62) shows that at the full-information optimum, the regulator should ensure that the marginal rate of substitution between the firm's utility and consumer surplus is the same in the two states. This is just an application of standard Ramsey principles.

Second-best optimum Now return to the setting where the regulator must motivate ϕ, and so the constraint $\phi = \hat{\phi}(\Delta^U)$ is relevant. Let $\hat{\lambda}$ denote the Lagrange multiplier associated with (54) in this second-best problem. Then the first-order conditions for the choice of U_i in expression (56) in this setting are

$$V_L'(U_L) = -(1+\hat{\lambda}) - \frac{\hat{\phi}'}{\hat{\phi}}\Delta^V; \qquad V_H'(U_H) = -(1+\hat{\lambda}) + \frac{\hat{\phi}'}{1-\hat{\phi}}\Delta^V, \tag{63}$$

where $\Delta^V \equiv V_L(U_L) - V_H(U_H)$ is the increment in consumer surplus in the desirable state L at the optimum. Notice that in the extreme case where the firm cannot affect the probability of a favorable outcome, so that $\hat{\phi}' = 0$, expression (63) collapses to the full-information condition in (62), and so the full-information outcome is attained.[79]

In the ensuing sections we consider the special cases of optimal regulation of a risk-neutral firm (case 1 in the preceding discussion) and a risk-averse firm (case 2). The discussion of the case of limited regulatory instruments (case 3) is deferred until Section 3.3.

2.6.1. *Regulation of a risk-neutral firm*

It is well known that when the firm is risk neutral and can bear unlimited losses ex post, the full-information outcome is attainable. To see why, substitute expression (57) into expected welfare (56). Doing so reveals that the regulator's objective is to maximize

$$W = \phi w_L\big(p_L^*\big) + (1-\phi) w_H\big(p_H^*\big) - D(\phi), \tag{64}$$

subject to $\phi = \hat{\phi}(\Delta^U)$ and the participation constraint (54). The regulator can structure the two utilities U_L and U_H to meet the firm's participation constraint (54) without affecting the firm's effort incentives. Since there is a one-to-one relationship between the incremental utility Δ^U and the effort level ϕ, the regulator will choose Δ^U to implement the value of ϕ that maximizes expression (64), and the full-information outcome is achieved.

PROPOSITION 8. *The full-information outcome is feasible* (*and optimal*) *in the pure moral hazard setting when the firm is risk-neutral and transfers can be employed. The optimal outcomes for the firm and for consumers are*

$$D'(\phi) = U_L - U_H = w_L\big(p_L^*\big) - w_H\big(p_H^*\big); \qquad V_L(U_L) = V_H(U_H). \tag{65}$$

The conclusion in Proposition 8 parallels the conclusion of the model of regulation under adverse selection when distributional concerns are absent (so $\alpha = 1$), discussed in Section 2.3.1. In both cases, the firm is made the residual claimant for the social surplus and consumers are indifferent about the realized state. In the present moral hazard setting, this requires that the firm face a high-powered incentive scheme. If state i occurs and the firm chooses price p_i, the regulator gives the firm a transfer payment

[79] The two multipliers λ and $\hat{\lambda}$ are equal in this case.

$T_i = v_i(p_i) - K$. Here, the constant K is chosen so that the firm makes zero rent in expectation. Under this policy, the firm has the correct incentives to set prices in each state, so $p_i = p_i^*$ is chosen. In addition, the firm has the correct incentives to choose ϕ to maximize social welfare in expression (64) because the firm has been made the residual claimant for the welfare generated by its actions.

2.6.2. Regulation of a risk-averse firm

When the relationship between firm utility and net consumer surplus is as specified in Equation (58), conditions (59) and (62) together imply that if the regulator could directly control the firm's effort ϕ, the outcomes for consumers and the firm would optimally be

$$U_L = U_H; \qquad V_L(U_L) - V_H(U_H) = w_L\big(p_L^*\big) - w_H\big(p_H^*\big). \tag{66}$$

In words, if the firm's effort could be controlled directly, the risk-averse firm should be given full insurance, so that it would receive the same utility (and rent) in each state. Of course, full insurance leaves the firm with no incentive to achieve the desirable outcome. In contrast, a higher-powered scheme ($U_L > U_H$) provides effort incentives, but leaves the firm exposed to risk.

The second-best policy is given by expression (63) above. In particular, it is still optimal to have the full-information prices p_i^* in each state i, since these prices maximize the available surplus that can be shared between the firm and consumers in any given state i.[80] Assuming that $w_L(p_L^*)$ is greater than $w_H(p_H^*)$, which is implied by the convention that L is the socially desirable state, expression (63) implies that[81]

$$U_L \geqslant U_H; \qquad V_L(U_L) \geqslant V_H(U_H). \tag{67}$$

Therefore, the firm is given an incentive to achieve the desirable outcome, but this incentive is sufficiently small that consumers are better off when the good state is realized. The more pronounced is the firm's aversion to risk, the more important is the need to insure the firm and the lower is the power of the optimal incentive scheme. In the limit as the firm becomes infinitely risk averse so that the firm's utility function in expression (54) becomes

$$U = \min\{R_L, R_H\} - D(\phi),$$

80 This is another version of the incentive-pricing dichotomy discussed in Laffont and Tirole (1993b): prices are employed to ensure allocative efficiency, while rent is employed to create incentives to increase productive efficiency.

81 To see this, suppose $\Delta^V \geqslant 0$ at the optimum. Then expression (63) implies $U_L \geqslant U_H$. Suppose by contrast $\Delta^V < 0$. Then expression (63) implies $U_L < U_H$. But since

$$\Delta^V = \big[w_L(p_L^*) - w_H\big(p_H^*\big)\big] - \big[u^{-1}(U_L) - u^{-1}(U_H)\big],$$

it follows that $\Delta^V > [w_L(p_L^*) - w_H(p_H^*)] > 0$ when $U_L < U_H$, which is a contradiction. Therefore, the only configuration consistent with expression (63) is $\Delta^V \geqslant 0$ and $U_L \geqslant U_H$, as claimed.

the firm does not respond to incentives since it cares only about its rent in the worse state. In this case, the firm delivers no effort to attain the desirable outcome, and the regulator does not benefit by setting $R_L > R_H$.

2.6.3. *Regulation of a risk-neutral firm with limited liability*

The analysis to this point has not considered any lower bounds that might be placed on the firm's returns. In practice, bankruptcy laws and liability limits can introduce such lower bounds. To analyze the effects of such bounds, the model of Section 2.6.1 is modified to incorporate an ex post participation constraint that the firm must receive rent $R_i \geqslant 0$ in each state. Since the firm now cannot be punished when there is a bad outcome, all incentives must be delivered through a reward when there is a good outcome.[82] In this case, the regulator will set $R_H = 0$ and use the rent in the good state to motivate the firm. The firm's overall expected rent is $\phi R_L - D(\phi)$, and it will choose effort ϕ to maximize this expression, so that $D'(\phi) = R_L$. Since the firm will enjoy positive expected rent in this model, the regulator's valuation of rent will be important for the analysis. Therefore, as with the adverse selection analysis, suppose the regulator places weight $\alpha \in [0, 1]$ on the firm's rent. In this case, much as in Section 2.6.1 above, the regulator's objective is to choose R_L to maximize

$$W = \phi\left\{w_L\left(p_L^*\right) - (1-\alpha)R_L\right\} + (1-\phi)w_H\left(p_H^*\right) - \alpha D(\phi),$$

subject to the incentive constraint $D'(\phi) = R_L$. (As before, it is optimal to set the full-information prices p_i^* and to use transfers to provide effort incentives.) Therefore, since $R_L = D'(\phi)$, the regulator should choose ϕ to maximize

$$W = \phi\left\{w_L\left(p_L^*\right) - (1-\alpha)D'(\phi)\right\} + (1-\phi)w_H\left(p_H^*\right) - \alpha D(\phi).$$

The solution to this problem has the first-order condition

$$D'(\phi) = w_L\left(p_L^*\right) - w_H\left(p_H^*\right) - (1-\alpha)\phi D''(\phi). \tag{68}$$

Comparing expression (68) with expression (65), the corresponding expression from the setting where there is no ex post participation constraint, it is apparent that this constraint produces lower equilibrium effort. (Recall $D''(\cdot) > 0$.) Therefore, the introduction of a binding limited liability constraint reduces the power of the optimal incentive scheme. The reduced power is optimal in the presence of limited liability because the regulator can no longer simply lower the firm's payoff when the unfavorable outcome is realized so as to offset any incremental reward that is promised when the favorable outcome is realized. The only situation where the power of the optimal incentive scheme is not reduced by the imposition of limited liability constraints is when the

[82] The ex ante participation constraint is assumed not to bind in the ensuing analysis. See Laffont and Martimort (2002, Section 3.5) for further discussion of limited liability constraints.

regulator has no strict preference for consumer surplus over firm rent ($\alpha = 1$), just as in the adverse selection paradigm.

Notice that this limited liability setting produces results similar to those obtained under risk aversion in Section 2.6.2. The full-information outcome is not feasible and too little effort is supplied relative to the full-information outcome in both settings. The limited liability setting also provides parallels with the adverse selection analysis in Section 2.3.1. In the limited liability setting, the regulator faces a trade-off between rent extraction and incentives. In the adverse selection settings, the regulator faces a corresponding trade-off.

2.6.4. Repeated moral hazard

Three primary additional considerations arise when the moral hazard model is repeated over time. First, the firm can effectively become less averse to risk, since it can pool the risk over time, and offset a bad outcome in one period by borrowing against the expectation of a good future outcome. Second, with repeated observations of the outcome, the regulator has better information about the firm's effort decisions (especially if current effort decisions have long-term effects). Third, the firm can choose from a wide range of possible dynamic strategies. For instance, the firm's managers can choose when to invest in effort, and might respond to a positive outcome in the current period by reducing effort in the future, for example. Consequently, the regulator's optimal inter-temporal policy, and the firm's profit-maximizing response to the policy, can be complicated.[83] In particular, the optimal policy typically will make the firm's reward for a good outcome in the current period depend on the entire history of outcomes, even in a setting where effort only affects the outcome in the current period. The dynamic moral hazard problem is discussed further in Section 3.2.3 below, where the optimal frequency of regulatory review is considered.

2.7. Conclusions

Asymmetric information about the regulated industry can greatly complicate the design of regulatory policy. This section has reviewed the central insights provided by the pioneering studies of this issue and by subsequent analyses. The review reveals that the manner in which the regulated firm is optimally induced to employ its superior knowledge in the best interests of consumers varies according to the nature of the firm's

[83] See Rogerson (1985). Holmstrom and Milgrom (1987) examine a continuous time framework in which the optimal inter-temporal incentive scheme is linear in the agent's total production. Bolton and Dewatripont (2005, ch. 11) emphasize the effects of limited commitment on the part of the principal. Also see the analyses of renegotiation by Fudenberg and Tirole (1990), Chiappori et al. (1994), Ma (1994), and Matthews (2001). Laffont and Martimort (2002, Section 8.2) analyze the two-period model with full commitment. Radner (1981, 1985) provides early work on the repeated moral hazard problem.

privileged information and according to the intertemporal commitment powers of the regulator.

The review emphasized five general principles. First, when a regulated firm has privileged information about the environment in which it operates, the firm typically is able to command rent from its superior information. Second, to help limit this rent, a regulator can design options from which the firm is permitted to choose. When designed appropriately, the options induce the firm to employ its superior industry knowledge to realize Pareto gains. Third, the options intentionally induce outcomes that differ from the outcomes the regulator would implement if he shared the firm's knowledge of the industry. These performance distortions serve to limit the firm's rent. Fourth, a benevolent regulator is better able to limit the firm's rent and secure gains for consumers via the careful design of regulatory options when he is endowed with a broader set of regulatory instruments and more extensive commitment powers. Fifth, when the regulator's commitment powers are limited, it may be optimal to limit the regulator's access to information, in order to limit inappropriate use of the information.

The analysis in this section has focused on the design of optimal regulatory policy when there is a single monopoly supplier of regulated services.[84,85] Section 4 reviews some of the additional considerations that arise in the presence of actual or potential competition. First, though, Section 3 discusses several simple regulatory policies, including some that are commonly employed in practice.

3. Practical regulatory policies

The discussion in Section 2 focused on optimal regulatory policy. Such analyses model formally the information asymmetry between the regulator and the firm and then determine precisely how the regulator optimally pursues his goals in the presence of this asymmetry. While this normative approach can provide useful insights for the design

[84] The analysis in this section also has taken as given the quality of the goods and services delivered by the regulated firm. Section 3 discusses policies that can promote increased service quality. Laffont and Tirole (1993b, ch. 4) and Lewis and Sappington (1992) discuss how regulated prices are optimally altered when they must serve both to motivate the delivery of high-quality products and to limit incentives to misrepresent private information. Lewis and Sappington (1991a) note that consumers and the regulated firm can both suffer when the level of realized service quality is not verifiable. In contrast, Dalen (1997) shows that in a dynamic setting where the regulator's commitment powers are limited, consumers may benefit when quality is not verifiable.

[85] The analysis in this section also has taken as given the nature of the information asymmetry between the regulator and the firm. Optimal regulatory policies will differ if, for example, the regulator wishes to motivate the firm to obtain better information about its environment, perhaps in order to inform future investment decisions. [See Lewis and Sappington (1997) and Crémer, Khalil and Rochet (1998a, 1998b), for example.] Iossa and Stroffolini (2002) show that optimal regulatory mechanisms of the type described in Proposition 3 provide the firm with stronger incentives for information acquisition than do price cap plans of the type considered in Section 3. Iossa and Stroffolini (2005) stress the merits of revenue sharing requirements under price cap regulation in this regard.

and evaluation of real-world regulatory policy, the approach has its limitations. In particular: (i) all relevant information asymmetries can be difficult to characterize precisely; (ii) the form of optimal regulatory policies is not generally known when information asymmetries are pronounced and multi-dimensional; (iii) a complete specification of all relevant constraints on the regulator and firm can be difficult to formulate; (iv) some of the instruments that are important in optimal reward structures (such as transfers) are not always available in practice; and (v) even the goals of regulators can be difficult to specify in some situations. Therefore, although formal models of optimal regulatory policy can provide useful insights about the properties of regulatory policies that may perform well in practice, these models seldom capture the full richness of the settings in which actual regulatory policies are implemented.[86]

This has led researchers and policy makers to propose relatively simple regulatory policies that appear to have some desirable properties, even if they are not optimal in any precise sense. The purpose of this section is to review some of these pragmatic policies. The policies are sorted on four dimensions: (1) the extent of pricing flexibility granted to the regulated firm; (2) the manner in which regulatory policy is implemented and revised over time; (3) the degree to which regulated prices are linked to realized costs; and (4) the discretion that regulators themselves have when they formulate policy. These dimensions are useful for expository purposes even though they incorporate substantial overlap.

To begin, it may be helpful to assess how two particularly familiar regulatory policies compare on these four dimensions. Table 27.1 provides a highly stylized interpretation of how price cap and rate-of-return regulation differ along these dimensions, and the broad effects of such policies. Under a common form of price cap regulation, the prices the firm charges for specified services are permitted to increase, on average, at a specified rate for a specified period of time. The specified average rate of price increase often is linked to the overall rate of price inflation, and typically does not reflect the firm's realized production costs or profit. In contrast, rate-of-return regulation specifies an allowed rate of return on investment for the firm, and adjusts the firm's prices as its costs change to ensure the firm a reasonable opportunity to earn the authorized return.

Table 27.1 reflects the idea that, at least under a common caricature of price cap regulation: (i) only the firm's average price is controlled (which leaves the firm free to control the pattern of relative prices within the basket of regulated services); (ii) the rate at which prices can increase over time is fixed for several years, and is not adjusted to reflect realized costs and profits during the time period; (iii) current prices are not explicitly linked to current costs; and (iv) the regulator has considerable discretion over future policy (once the current price control period has expired). Because prices are not directly linked to costs for relatively long periods of time, the firm can have strong incentives to reduce its operating costs. By contrast, under the classic depiction of rate-of-return regulation: (i) the regulator sets prices, and affords the firm little discretion in

[86] See Crew and Kleindorfer (2002) and Vogelsang (2002) for critical views regarding the practical relevance of the recent optimal regulation literature.

Table 27.1
Price cap versus rate-of-return regulation

	Price cap	Rate-of-return
Firm's flexibility over relative prices	Yes	No
Regulatory lag	Long	Short
Sensitivity of prices to realized costs	Low	High
Regulatory discretion	Substantial	Limited
Incentives for cost reduction	Strong	Limited
Incentives for durable sunk investment	Limited	Strong

altering these prices; (ii) prices are adjusted as necessary to ensure that the realized rate of return on investment does not deviate substantially from the target rate; (iii) prices are adjusted to reflect significant changes in costs; and (iv) the regulator is required to ensure that the firm has the opportunity to earn the target rate of return on an ongoing basis. Because the firm is ensured a reasonable opportunity to earn the authorized return on its investments over the long term, the firm has limited concern that its sunk investments will be expropriated by future regulatory policy.

Rate-of-return and price cap regulation can have different effects on unobservable investment (e.g., managerial effort) designed to reduce operating costs and observable infrastructure investment. Because it links prices directly to realized costs, rate-of-return regulation is unlikely to induce substantial unobserved cost-reducing investment. However, rate-of-return regulation can promote observable infrastructure investment by limiting the risk that such investment will be expropriated. In contrast, price cap regulation can provide strong incentives for unobservable cost-reducing effort, especially when the regulatory commitment period (the length of time between regulatory reviews) is relatively long. Therefore, the choice between these two forms of regulation will depend in part on the relative importance of the two forms of investment. In settings where the priority is to induce the regulated firm to employ its existing infrastructure more efficiently, a price cap regime may be preferable. In settings where it is important to reverse a history of chronic under-investment in key infrastructure, a guaranteed rate of return on (prudently incurred) investments may be preferable.[87] The relative performance of price cap and rate-of-return regulation is explored in more detail and in other dimensions in the remainder of Section 3.[88]

[87] Regulatory regimes also differ according to the incentives they provide the firm to modernize its operating technology. In contrast to rate-of-return regulation, for example, price cap regulation can encourage the regulated firm to replace older high-cost technology with newer low-cost technology in a timely fashion. It can do so by severing the link between the firm's authorized earnings and the size of its rate base. See Biglaiser and Riordan (2000) for an analysis of this issue.

[88] For more detailed discussions of the key differences between price cap regulation and rate-of-return regulation, see, for example, Acton and Vogelsang (1989), Cabral and Riordan (1989), Hillman and Braeutigam (1989), Clemenz (1991), Braeutigam and Panzar (1993), Liston (1993), Armstrong, Cowan and Vickers

3.1. Pricing flexibility

In a setting where the regulated firm has no privileged information about its operating environment, there is little reason for the regulator to delegate pricing decisions to the firm. Such delegation would simply invite the firm to implement prices other than those that are most preferred by the regulator. In contrast, if the firm is better informed than the regulator about its costs or about consumer demand, then, by granting the firm some authority to set its tariffs, the regulator may be able to induce the firm to employ its superior information to implement prices that generate higher levels of welfare than the regulator could secure by dictating prices based upon his limited information. A formal analysis of this possibility is presented in Section 3.1.1. Section 3.1.2 compares the merits of two particular means by which the firm might be afforded some flexibility over its prices: *average revenue* regulation and *tariff basket* regulation.

Despite the potential merits of delegating some pricing flexibility to the regulated firm, there are reasons why a regulator might wish to limit the firm's pricing discretion. One reason is that the regulated firm may set prices to disadvantage rivals, as explained in Section 3.1.3. A second reason is the desire to maintain pricing structures that reflect distributional or other political objectives. In practice, regulators often limit a firm's pricing flexibility in order to prevent the firm from unraveling price structures that have been implemented to promote social goals such as universal service (i.e., ensuring that nearly all citizens consume the service in question).

3.1.1. The cost and benefits of flexibility with asymmetric information

The merits of affording the regulated firm some discretion in setting prices vary according to whether the firm is privately informed about its costs or its demand.[89] We assume that transfer payments to or from the firm are not feasible, and the firm's tariff must be designed to cover its costs. As in Section 2, the regulator seeks to maximize a weighted average of expected consumer surplus and profit, where $\alpha \leqslant 1$ is the weight the regulator places on profit.

Asymmetric cost information Suppose first that the firm has superior knowledge of its (exogenous) cost structure, while the regulator and firm are both perfectly informed about industry demand. The regulated firm produces n products. The price for product i is p_i, and the vector of prices that the firm charges for its n products is $\mathbf{p} = (p_1, \ldots, p_n)$. Suppose that consumer surplus with prices $\mathbf{p}$ is $v(\mathbf{p})$, where this function is known to all parties. Suppose also that the firm's total profit with prices $\mathbf{p}$ is $\pi(\mathbf{p})$. Since the firm has superior information about its costs in this setting, the regulator is not completely informed about the firm's profit function, $\pi(\cdot)$.

(1994), Blackmon (1994), Mansell and Church (1995), Sappington (1994, 2002), Sappington and Weisman (1996a), and Joskow (2005).

[89] This discussion is based on Armstrong and Vickers (2000).

Some pricing flexibility is always advantageous in this setting. To see why, suppose the regulator instructs the firm to offer the fixed price vector $\mathbf{p}^0 = (p_1^0, \dots, p_n^0)$. Provided these prices allow the firm to break even, so that the firm agrees to participate, this policy yields welfare $v(\mathbf{p}^0) + \alpha\pi(\mathbf{p}^0)$. Suppose instead, the regulator allows the firm to choose any price vector that leaves consumers in aggregate just as well off as they were with $\mathbf{p}^0$, so that the firm can choose any price vector

$$\mathbf{p} \in \mathcal{P} = \{\mathbf{p} \mid v(\mathbf{p}) \geqslant v(\mathbf{p}^0)\}. \tag{69}$$

By construction, this regulatory policy ensures that consumers are no worse off in aggregate than they are under the fixed pricing policy $\mathbf{p}^0$.[90] Furthermore, the firm will be strictly better off when it can choose a price from the set $\mathcal{P}$, except in the knife-edge case where $\mathbf{p}^0$ happens to be the most profitable way to generate consumer surplus $v(\mathbf{p}^0)$. Therefore, welfare is sure to increase when the firm is granted pricing flexibility in this manner.[91]

Asymmetric demand information The merits of pricing flexibility are less clear cut when the firm has superior knowledge of industry demand. To see why it might be optimal not to grant the firm any authority to set prices when consumer demand is private information, suppose the firm has known, constant marginal costs $\mathbf{c} = \{c_1, \dots, c_n\}$ for its n products and has no fixed cost of operation. Then, regardless of the form of consumer demand, the full-information outcome is achieved by constraining the firm to offer the single price vector $\mathbf{p} = \mathbf{c}$, so that prices are equal to marginal costs. If the firm is given the flexibility to choose from a wider set of price vectors, it will typically choose prices that deviate from costs, thereby reducing welfare.

More generally, whether the firm should be afforded any pricing flexibility depends on whether the full-information prices are incentive compatible. In many natural cases, a firm will find it profitable to raise price when demand increases. However, welfare considerations suggest that prices should be higher in those markets with relatively inelastic demand, not necessarily in markets with large demand. Thus, if an increase in demand is associated with an increase in the demand elasticity, the firm's incentives are not aligned with the welfare-maximizing policy, and so it is optimal to restrict the firm to offer a single price vector. If, by contrast, an increase in demand is associated with a reduction in the market elasticity, then private and social incentives coincide, and it is optimal to afford the firm some authority to set prices.

This analysis parallels the analysis in Section 2.3.2 of the optimal regulation (with transfers) of a single-product firm that is privately informed about its demand function. In that setting, when the firm has a concave cost function, an increase in demand is

[90] Since some prices will increase under this policy, some individual consumers may be made worse off.

[91] Notice that the profit-maximizing prices for the firm operating under this constraint are closely related to Ramsey prices: profit is maximized subject to a constraint on the level of consumer surplus achieved, or, equivalently, consumer surplus is maximized subject to a profit constraint. However, the prices are not true Ramsey prices since the firm's rent will not be zero in general.

associated with a lower marginal cost. Therefore, the firm's incentives – which typically are to set a higher price in response to greater demand – run counter to social incentives, which are to set a lower price when marginal cost is lower, i.e., when demand is greater. These conflicting incentives make it optimal to give the firm no authority to choose its prices.

In summary, unequivocal conclusions about the merits of granting pricing flexibility to a regulated firm are not available. In practice, a regulated firm typically will be better informed than the regulator about both its demand and its cost structure. Furthermore, the regulator will often be unaware of the exact nature of likely variation in demand. Consequently, the benefits that pricing flexibility will secure in any specific setting may be difficult to predict in advance. However, the foregoing principles can inform the choice of the degree of pricing flexibility afforded the firm.

3.1.2. Forms of price flexibility

The merits of affording the regulated firm some pricing flexibility vary with the form of the contemplated flexibility. To illustrate this point, consider two common variants of average price regulation: average revenue regulation and tariff basket regulation.[92] Suppose consumer demand for product i with the price vector $\mathbf{p}$ is $Q_i(\mathbf{p})$, and $v(\mathbf{p})$ is the corresponding total consumer surplus function. In order to compare outcomes under various regimes, notice that, for any pair of price vectors $\mathbf{p}^1$ and $\mathbf{p}^2$, the following inequality holds[93]

$$v(\mathbf{p}^2) \geqslant v(\mathbf{p}^1) - \sum_{i=1}^{n} (p_i^2 - p_i^1) Q_i(\mathbf{p}^1). \tag{70}$$

Expression (70) states that consumer surplus with price vector $\mathbf{p}^2$ is at least as great as consumer surplus with price vector $\mathbf{p}^1$, less the difference in revenue generated by the two price vectors when demands are $Q_i(\mathbf{p}^1)$. The inequality follows from the convexity of the consumer surplus function.

Average revenue regulation In its simplest form, average revenue regulation limits to a specified level, $\bar{p}$, the average revenue the firm derives from its regulated operations. Formally, the average revenue constraint requires the firm's price vector to lie in the set

$$\mathbf{p} \in \mathcal{P}^{AR} = \left\{ \mathbf{p} \;\middle|\; \frac{\sum_{i=1}^{n} p_i Q_i(\mathbf{p})}{\sum_{i=1}^{n} Q_i(\mathbf{p})} \leqslant \bar{p} \right\}. \tag{71}$$

[92] This section is based on Armstrong and Vickers (1991).

[93] The expression to the right of the inequality in (70) reflects the level of consumer surplus that would arise under prices $\mathbf{p}^1$ if consumers did not alter their consumption when prices changed from $\mathbf{p}^2$ to $\mathbf{p}^1$ (and instead just benefited from the monetary savings permitted by the new prices). Since consumers generally will be able to secure more surplus by altering their consumption in response to new prices, the inequality follows.

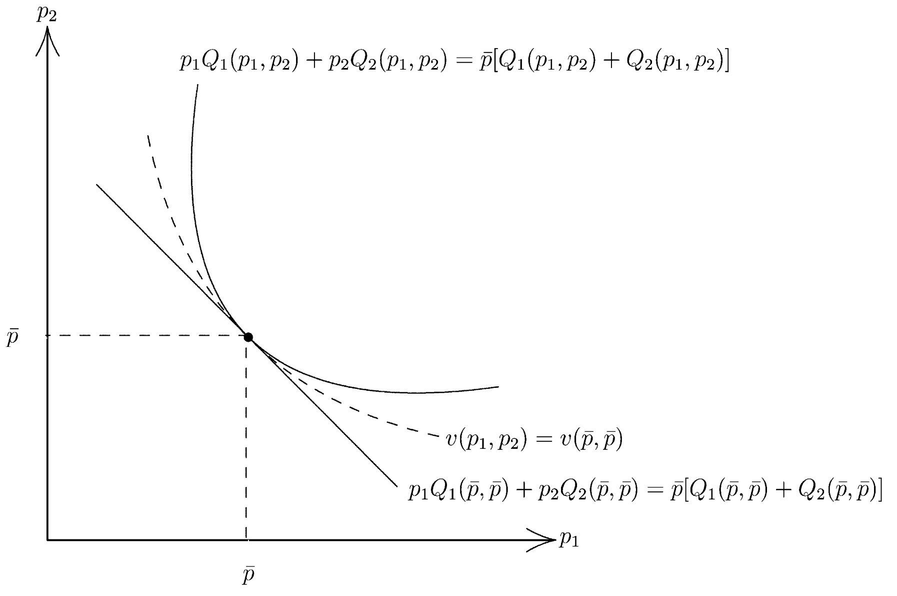

Figure 27.2. Average revenue and tariff basket regulation.

The term to the left of the inequality in expression (71) is average revenue: total revenue divided by total output.[94] Notice that if $\mathbf{p}^2$ is the vector of prices where all services have the same benchmark price, $\bar{p}$, and $\mathbf{p}^1$ is any price vector that satisfies the average revenue constraint in (71) exactly, then inequality (70) implies that $v(\mathbf{p}^1) \leqslant v(\mathbf{p}^2)$. Therefore, regardless of the prices the firm sets under this form of regulation, consumers will be worse off than if the firm were required to set price $\bar{p}$ for each of its products.[95] The reduction in consumer surplus arises because as the firm raises prices, the quantity demanded decreases, which reduces average revenue, and thereby relaxes the average revenue constraint.

This reduction in consumer surplus is illustrated in Figure 27.2 for the case where the firm offers two products. Here the boundary of the set $\mathcal{P}^{AR}$ in (71) lies inside the set of price vectors that make consumers worse off than they are with the uniform price vector $(\bar{p}, \bar{p})$. (Consumer surplus declines with movements to the north-east in Figure 27.2.)

The following result summarizes the main features of average revenue regulation:

[94] Since total output is calculated by summing individual output levels, average revenue regulation in this form is most appropriate in settings where the units of output of the n regulated products are commensurate.

[95] Armstrong, Cowan and Vickers (1995) show that, for similar reasons, allowing non-linear pricing reduces consumer surplus when average revenue regulation is imposed on the regulated firm, compared to a regime where the firm offers a linear tariff.

PROPOSITION 9. (i) *Consumer surplus is lower under binding average revenue regulation when the firm is permitted to set any prices that satisfy inequality* (71) *rather than being required to set each price at* $\bar{p}$.

(ii) *Total welfare* (*the weighted sum of consumer surplus and profit*) *could be higher or lower when the firm is permitted to set any prices that satisfy inequality* (71) *rather than being required to set each price at* $\bar{p}$.

(iii) *Consumer surplus can decrease under average revenue regulation when the authorized level of average revenue* $\bar{p}$ *declines.*

Part (ii) of Proposition 9 states that, although consumers are necessarily worse off with average revenue regulation, the effect on total welfare is ambiguous because the pricing discretion afforded the firm leads to increased profit. This increased profit might outweigh the reduction in consumer surplus. Part (iii) of Proposition 9 indicates that a more stringent price constraint is not always in the interests of consumers under average revenue regulation. To see why, consider the firm's incentives as the authorized level of average revenue $\bar{p}$ declines. Clearly, average revenue, as calculated in expression (71), does not vary with production costs. Consequently, a required reduction in average revenue may be achieved with the smallest reduction in profit by reducing the sales of those products that are particularly costly to produce. If consumers value these products highly, then the reduction in consumer welfare due to the reduced consumption of highly-valued products can outweigh any increase in consumer welfare due to the reduction in average prices that accompanies a reduction in $\bar{p}$.[96]

The drawbacks of average revenue regulation can be illustrated in the case where the regulated firm sells a single product using a two-part tariff. This tariff consists of a fixed charge A and a per-unit price p. Suppose the firm is required to keep calculated average revenue below a specified level $\bar{p}$. Then, as long as the number of consumers is invariant to the firm's pricing policy over the relevant range of prices, the regulatory constraint (71) is

$$p + \frac{A}{Q(p)} \leqslant \bar{p}. \tag{72}$$

Inequality (72) makes apparent the type of strategic pricing that could be profitable for the firm under average revenue regulation. By setting a low usage price p, the firm can induce consumers to purchase more of its product. The increased consumption enables the firm to set a higher fixed charge without violating the average revenue constraint. From Proposition 9, this strategic pricing always causes consumer surplus to

[96] See Bradley and Price (1988), Law (1995), and Cowan (1997b), for example. Flores (2005) identifies conditions under which a more stringent price cap constraint can increase a firm's profit by allowing it to credibly commit to a more aggressive pricing, and thereby induce accommodating actions from rivals. Kang, Weisman and Zhang (2000) demonstrate that the impact of a tighter price cap constraint on consumer welfare can vary according to whether the basket of regulated services contains independent, complementary, or substitute products.

fall compared to the case where the firm is required to charge $\bar{p}$ for each unit of output (and set $A = 0$). Moreover, aggregate welfare may fall when two-part pricing is introduced under an average revenue constraint.[97] The profit-maximizing behavior of the firm under the average revenue constraint in inequality (72) is readily calculated in the setting where consumer participation in the market is totally inelastic and the firm has a constant marginal cost c per unit of supply. Since the firm's profit is increasing in A, the average revenue constraint (72) will bind, and so the firm's profit is

$$\pi = (p - c)Q(p) + A = (\bar{p} - c)Q(p).$$

Therefore, assuming $\bar{p} > c$ (as is required for the firm to break even), the firm sets its unit price p to maximize output, so that p is chosen to be as small as possible.[98] Consequently, average revenue regulation in this setting induces a distorted pattern of demand: the unit price is too low (below cost), while consumers pay a large fixed charge (a combination that makes consumers worse off compared to the case where they pay a constant linear price $\bar{p}$). In effect, under average revenue regulation, the firm effectively is allowed a margin $\bar{p} - c$ per unit of its output, and so it has an incentive to expand output inefficiently.[99]

Tariff basket regulation Tariff basket regulation provides an alternative means of controlling the overall level of prices charged by a regulated firm while affording the firm pricing flexibility. One representation of tariff basket regulation specifies reference prices, $\mathbf{p}^0$, and permits the firm to offer any prices that would reduce what consumers have to pay for their preferred consumption at the reference prices $\mathbf{p}^0$. Formally, the firm must choose prices that lie in the set

$$\mathbf{p} \in \mathcal{P}^{TB} = \left\{ \mathbf{p} \;\middle|\; \sum_{i=1}^{n} p_i Q_i(\mathbf{p}^0) \leqslant \sum_{i=1}^{n} p_i^0 Q_i(\mathbf{p}^0) \right\}. \tag{73}$$

Under this form of tariff basket regulation, the weights that are employed to calculate the firm's average price are exogenous to the firm, and are proportional to consumer demands at the reference prices $\mathbf{p}^0$.

Notice that consumers are always better off with this form of regulation than they would be with the reference tariff $\mathbf{p}^0$.[100] This form of tariff basket regulation is a linear

[97] See Sappington and Sibley (1992), Cowan (1997a) and Currier (2005) for dynamic analyses along these lines. The firm's ability to manipulate price cap constraints can be limited by requiring the firm to offer the uniform tariff $(p^0, 0)$ each year in addition to any other tariff (p, A) that satisfies the price cap constraint – see Vogelsang (1990), Sappington and Sibley (1992), and Armstrong, Cowan and Vickers (1995).

[98] That is to say, the price is zero if a zero price results in finite demand.

[99] This conclusion is similar to Averch and Johnson's (1962) finding regarding over-investment under rate-of-return regulation. In their model, the regulated firm earns a return on capital that exceeds the cost of capital. Consequently, the firm employs more than the cost-minimizing level of capital.

[100] This follows from inequality (70) if we let $\mathbf{p}^1$ be the reference price vector $\mathbf{p}^0$ and let $\mathbf{p}^2$ be any vector in the set $\mathcal{P}^{TB}$ defined in expression (73).

approximation to the regulatory policy specified in expression (69). In particular, the set of prices in (73) lies inside the set (69) which, by construction, is the set of prices that make consumers better off than they are with prices $\mathbf{p}^0$. This finding is illustrated in Figure 27.2 for the case where the reference price vector $\mathbf{p}^0$ is $(\bar{p}, \bar{p})$. The boundary of the region of feasible prices $\mathcal{P}^{TB}$ in expression (73) is the straight line in the figure. Since this line lies everywhere below the locus of prices at which consumer surplus is $v(\bar{p}, \bar{p})$, consumers are better off when the regulated firm is given the pricing flexibility reflected in expression (73). Since the firm will also be better off with the flexibility permitted in constraint (73), it follows that welfare is higher under this form of regulation than under the fixed price vector $\mathbf{p}^0$.

The benefits of this form of regulation are evident in the case where the regulated firm sets a two-part tariff, with fixed charge A and unit price p, for the single product it sells. Here, the reference tariff is just the linear tariff where each unit of the product is sold at price p^0. In this case, constraint (73) becomes

$$A + pQ\big(p^0\big) \leqslant p^0 Q\big(p^0\big).$$

Assuming that consumer participation does not vary with the established prices, this constraint will bind, and so the firm's per-consumer profit with the unit price p is

$$\pi = \big(p^0 - p\big)Q\big(p^0\big) + (p - c)Q(p),$$

where c is the firm's constant marginal cost of production. It is readily shown that the profit-maximizing price p lies between the reference price and cost: $c < p < p^0$. This outcome generates more consumer surplus and higher welfare than does the linear price p^0.

Although this form of tariff basket regulation can secure increased consumer surplus and welfare, its implementation requires knowledge of demands at the reference prices $\mathbf{p}^0$ even when those prices are not actually chosen. Thus, demand functions must be known in static settings. By contrast, with average revenue regulation – where the weights in the price index reflect actual, not hypothetical, demands – only realized demands at the actual prices offered need to be observed. In dynamic settings, outputs in the previous period might be employed as current period weights when implementing tariff basket regulation, as explained in Section 3.2.1 below.

3.1.3. Price flexibility and entry

The type of pricing flexibility afforded the regulated firm can have important effects on the firm's response to entry by competitors.[101] To illustrate this fact, suppose the incumbent firm operates in two separate markets. Suppose further that if entry occurs at all, it will occur in only one of these markets. There are then four natural pricing regimes to consider:

[101] This discussion is based on Armstrong and Vickers (1993). See Anton, Vander Weide and Vettas (2002) for further analysis.

1. *Laissez-faire*: Here the incumbent can set its preferred prices in the two markets.
2. *Ban on price discrimination*: Here the incumbent can set any prices it desires, as long as the prices are the same in the two markets. (Regulators often implement such policies with the stated aim of bringing the benefits of competition to all consumers, including those who reside in regions where direct competition among firms is limited.) Here, if the incumbent lowers its price in one market in response to entry, it must also lower its price in the other market, even if entry is not a threat in that market.
3. *Separate price caps*: Here the incumbent faces a distinct upper limit on the price it can charge in each market. Because there is a distinct price cap in each market, the price the firm sets in one market does not affect the price it can charge in the other market.
4. *Average price cap*: Here the incumbent operates under a single price cap that limits the average price charged in the two markets. Under such an average price cap, if the incumbent lowers its price in one market in response to entry, it can raise its price in the other market without violating the average price cap. Thus, in contrast to the case where price discrimination is not permitted, feasible prices have an inverse relationship here.

Regimes 1 and 2 here apply to situations where the firm is unregulated, at least in terms of the level of its average tariff, whereas regimes 3 and 4 entail explicit regulation of price levels.

These four policies will induce different responses to entry by the incumbent supplier. To illustrate this fact, suppose there is a sunk cost of entry, so the potential entrant will only enter if it anticipates profit in excess of this sunk cost. Once entry takes place, some competitive interaction occurs. Under regime 2, which bans price discrimination, the incumbent will tend to accommodate entry. This is because any price reduction in the competitive market forces the incumbent to implement the same price reduction in the captive market, which can reduce the incumbent's profit in the captive market. The incumbent's resulting reluctance to cut prices in response to entry can result in higher profit for the entrant. Thus, a restriction on the regulated firm's pricing discretion can act as a powerful form of entry assistance. In particular, a ban on price discrimination can induce entry that would not occur under the *laissez-faire* regime, which, in turn, can cause prices in both markets to fall below their levels in the *laissez-faire* regime.

The average price cap regime induces the opposite effects. The incumbent will react more aggressively to entry under an average price cap regime than under a regime that imposes a separate cap in each market. In particular, the incumbent may reduce the price it charges in the competitive market below its marginal cost because of the high price it can then charge in the captive market. Therefore, an average price cap regime can act as a powerful source of entry deterrence. This observation implies the merits of granting the firm some authority to set its prices – for instance, by regulating the firm under an average price cap instead of separate caps – require careful study when entry is a possibility. This issue is analyzed further in Section 5.2, which considers the regulation of a vertically-integrated supplier.

3.2. Dynamics

Regulatory policies also vary according to their implementation over time. A regulatory policy may be unable to secure substantial surplus for consumers when it is first implemented, but repeated application of the policy may serve consumers well. This section provides a four-part discussion of dynamic elements of regulatory policy. Section 3.2.1 considers different forms of dynamic average price regulation when transfer payments from the regulator to the firm are not permitted. Section 3.2.2 extends the analysis to allow the regulator to make transfers to the firm. Section 3.2.3 examines how frequently a firm's prices should be realigned to match its observed costs. Section 3.2.4 discusses the effect of (exogenous) technological change on prices.

3.2.1. Non-Bayesian price adjustment mechanisms: no transfers

First consider the natural dynamic extension of the tariff basket form of price regulation analyzed in Section 3.1.2. In this dynamic extension, the weights employed in the current regulatory period reflect the previous period's outputs.[102] Call the initial period in this dynamic setting period 0, and label subsequent periods $t = 1, 2, \ldots$. Let $\mathbf{p}^t = (p_1^t, \ldots, p_n^t)$ denote the vector of prices the firm charges for its n regulated products in period t. Let $\mathbf{q}^t = (q_1^t, \ldots, q_n^t)$ denote the corresponding vector of outputs, where $q_i^t = Q_i(\mathbf{p}^t)$. Tariff basket regulation in this dynamic setting states that if the price vector was $\mathbf{p}^{t-1}$ in period $t-1$, the firm can choose any price vector $\mathbf{p}^t$ in period t satisfying

$$\mathbf{p}^t \in \mathcal{P}^t = \left\{ \mathbf{p}^t \;\middle|\; \sum_{i=1}^{n} p_i^t q_i^{t-1} \leqslant \sum_{i=1}^{n} p_i^{t-1} q_i^{t-1} \right\}. \tag{74}$$

For now, assume the initial price vector $\mathbf{p}^0$ is specified exogenously. (This assumption will be revisited shortly.) Notice that the regulator only needs to observe the firm's (lagged) realized sales in order to implement this regulatory policy. In contrast, to implement the static version of tariff basket regulation considered in Section 3.1.2, the regulator needed to know the demand functions themselves (since he needed to know demands at the reference prices $\mathbf{p}^0$). Note that expression (74) can be written as

$$\mathbf{p}^t \in \mathcal{P}^t = \left\{ \mathbf{p}^t \;\middle|\; \sum_{i=1}^{n} \frac{R_i^{t-1}}{R^{t-1}} \left[\frac{p_i^t - p_i^{t-1}}{p_i^{t-1}} \right] \leqslant 0 \right\}, \tag{75}$$

where $R_i^{t-1} = p_i^{t-1} q_i^{t-1}$ is the revenue generated by product i in period $t-1$, and R^{t-1} is total revenue from the n products in period $t-1$. Constraint (75) states that a weighted average of proportional price increases cannot be positive in any period, where the weights are revenue shares in the preceding period.

[102] This discussion is based on Vogelsang (1989).

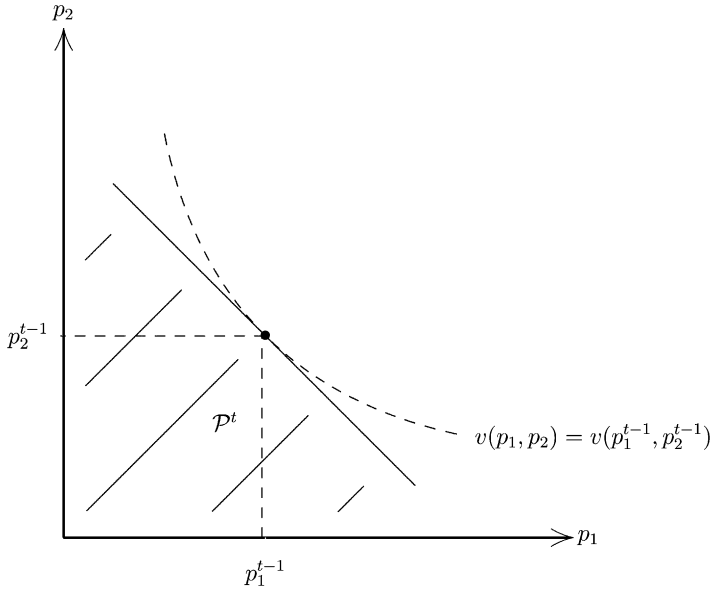

Figure 27.3. Dynamic tariff basket regulation.

Figure 27.3 illustrates how this form of dynamic average price regulation evolves over time. For the reasons explained in Section 3.1.2, any price vector in the set (74) generates at least the level of consumer surplus generated in the previous period, so $v(\mathbf{p}^t) \geqslant v(\mathbf{p}^{t-1})$. In particular, compared to the regime where the firm is forced to charge the same price vector $\mathbf{p}^0$ in each period, this more flexible regime yields higher welfare: consumers are better off (in each period) and, since the firm can implement the same vector $\mathbf{p}^0$ in each period if it chooses to do so, the firm is also better off. This dynamic process converges and the steady-state price vector will have the Ramsey form: profit is maximized subject to consumer surplus exceeding some specified level.[103] However, as in Section 3.1.1, long-run prices may diverge from Ramsey prices because the firm's rent is not necessarily zero.

The regulator might choose the initial price vector $\mathbf{p}^0$ to ensure that the firm makes only small rent in the long term and that total discounted expected welfare is maximized. Such a calculation would require considerable information, however. Alternatively, the firm might set $\mathbf{p}^0$ without constraint, but knowing that it will subsequently be controlled by the regulatory mechanism (74). In this setting, the firm will set its initial prices strategically in order to affect the weights in future constraints. For instance, the firm

[103] In a steady state, the firm's (short-run) profit-maximizing price vector in period t, $\mathbf{p}^t$, must be the same as the previous period's price vector, $\mathbf{p}^{t-1}$. From Figure 27.3, this implies that the firm's profit contour is tangent to the consumer surplus contour.

can set a high price for product i in period 0, and thereby reduce the weight applied to the price of product i in period 1. The net effect of such strategic pricing can be to reduce aggregate welfare below the level achieved in the absence of any regulation.[104]

Tariff basket regulation can also invite strategic pricing distortions when consumer demand and/or production costs are changing over time in predictable ways. To illustrate, the regulated firm will typically find it profitable to raise the price of a product for which consumer demand is increasing over time. Lagged output levels understate the actual losses a price increase imposes on consumers when demand is increasing over time. In this sense, tariff basket regulation does not penalize the firm sufficiently for raising prices on products for which demand is growing, and so induces relatively high prices on these products.[105]

Although this form of dynamic regulation leads to an increase in consumer surplus in every period, it need not lead to a particularly high level of consumer surplus. In particular, the firm may continue to make positive rent in the long run, even if the environment is stationary. One possible way to mitigate this problem, especially when demand is growing exogenously or when costs are falling exogenously, is to require average price reductions over time, so that average prices are required to fall proportionally by a factor X, say, in each period.[106] Formally, constraint (75) is modified to

$$\mathbf{p}^t \in \mathcal{P}^t = \left\{\mathbf{p}^t \;\middle|\; \sum_{i=1}^{n} \frac{R_i^{t-1}}{R^{t-1}}\left[\frac{p_i^t - p_i^{t-1}}{p_i^{t-1}}\right] \leqslant -X\right\}. \tag{76}$$

The key difficulty in implementing this mechanism, of course, is the choice of X. If X is too small (compared to potential productivity gains), the firm may be afforded substantial, persistent rent. In contrast, if X is too large, the firm may encounter financial difficulties. In a stationary environment, any positive value of X will eventually cause the firm to incur losses.

One possible way to determine an appropriate value for X involves the use of historic data on the firm's expenditures. To illustrate this approach, consider a policy that allows the regulated firm to set any price vector for its products in a given period, as long as the prices generate non-positive accounting profit for the firm when applied to outputs and costs in the previous period.[107] Suppose the firm's observable production expenditures in year t are E^t.[108] Formally, this policy permits the firm to select any vector of prices

[104] See Law (1997). Foreman (1995) identifies conditions under which strategic pricing to relax the price cap constraint is more pronounced when relative revenue weights are employed than when quantity weights are employed.

[105] Brennan (1989), Neu (1993), and Fraser (1995) develop this and related observations.

[106] We will discuss other aspects of this issue in Section 3.2.4.

[107] This policy is proposed and analyzed in Vogelsang and Finsinger (1979).

[108] For simplicity, we abstract from intertemporal cost effects, so that all costs of producing output in period t are incurred in period t.

in period t that lie in the set

$$\mathbf{p}^t \in \mathcal{P}^t = \left\{ \mathbf{p} \;\middle|\; \sum_{i=1}^{n} p_i q_i^{t-1} \leqslant E^{t-1} \right\}. \tag{77}$$

This policy, which we term the *lagged expenditure* policy, differs from the regulatory regime reflected in expression (74) in that last period's expenditure replaces last period's revenue as the cap on the current level of calculated revenue. Letting $\Pi^t = \sum_{i=1}^n p_i^t q_i^t - E^t$ denote the firm's observed profit in period t, constraint (77) can be re-written as

$$\mathbf{p}^t \in \mathcal{P}^t = \left\{ \mathbf{p} \;\middle|\; \sum_{i=1}^{n} p_i q_i^{t-1} \leqslant \sum_{i=1}^{n} p_i^{t-1} q_i^{t-1} - \Pi^{t-1} \right\}.$$

Thus, prices in each period must be such that the amount consumers would have to pay for the bundle of regulated products purchased in the preceding period decreases sufficiently to eliminate the observed profit of the firm in the previous period (and does not simply decrease, as in expression (74)). Expression (70) reveals that $v(\mathbf{p}^t) \geqslant v(\mathbf{p}^{t-1}) + \Pi^{t-1}$. Therefore, any profit the firm enjoys in one period is (more than) transferred to consumers in the next period. Notice that the regulator only needs to observe the firm's realized revenues and costs in order to implement the mechanism. The regulator does not need to know the functional form of the demand or cost functions in the industry.

Even though it can be implemented with very little information, the lagged expenditure policy can induce desirable outcomes under certain (stringent) conditions. In particular, the mechanism can sometimes eventually induce Ramsey prices (i.e., the prices that maximize surplus while securing non-negative rent for the firm). This conclusion is summarized in Proposition 10.

PROPOSITION 10. *Suppose demand and cost functions do not change over time and the firm's technology exhibits decreasing ray average cost.*[109] *Further suppose the regulated firm maximizes profit myopically each period. Then the lagged expenditure policy induces the firm to set prices that converge to Ramsey prices.*

The conditions under which the policy secures Ramsey prices are restrictive. If demand or cost functions shift over time, convergence is not guaranteed, and the regulated firm may experience financial distress. Even in a stationary environment, the non-myopic firm can delay convergence to the Ramsey optimum and reduce welfare substantially in the process. It can do so, for example, by intentionally increasing production costs above their minimum level. This behavior reflects the general proposition

[109] The cost function $C(\mathbf{q})$ exhibits decreasing ray average cost if $rC(\mathbf{q}) \geqslant C(r\mathbf{q})$ for all $r \geqslant 1$ and output vectors $\mathbf{q}$. Decreasing ray average costs ensure the firm can continue to secure non-negative profit under the mechanism as prices decline and outputs increase.

that when the firm's (current or future) permitted prices increase as the firm's current realized costs increase, the firm has limited incentives to control these costs.

To illustrate this last point, suppose the firm produces a single product and has a constant unit cost in each period, which the regulator can observe. If unit cost is c^{t-1} in the previous period, then the policy in expression (77) requires the firm to set a price no higher than c^{t-1} in the current period. Suppose that the firm can simply choose the unit cost, subject only to the constraint that $c^t \geqslant c$, where c is the firm's true (minimum possible) unit cost. Thus, any choice $c^t > c$ constitutes "pure waste". (Note that this inflated cost is actually incurred by the firm, and not simply misreported.) The firm discounts future profit at the rate δ, and its discounted profit in period zero is $\sum_{t=0}^{\infty} \delta^t Q(p^t)(p^t - c^t)$. The regulator chooses the initial price $p^0 > c$, and subsequently follows the rule $p^t = c^{t-1}$. If there were no scope for pure waste, the observed unit cost in period 0 would be c, and the firm would make profit $Q(p^0)(p^0 - c)$ for one period. It would make no profit thereafter, because price would equal unit cost in all subsequent periods. However, when δ is sufficiently large, the firm can increase the present discounted value of its profit by undertaking pure waste. To see why, notice that the firm could choose an inflated cost $c_H > c$ in period 0, and then implement the minimum cost c in every period thereafter. With this strategy, the firm's discounted profit is

$$Q\big(p^0\big)\big(p^0 - c_H\big) + \delta Q(c_H)(c_H - c). \tag{78}$$

Expression (78) is increasing in c_H at $c_H = c$ when $\delta Q(c) > Q(p^0)$. Consequently, whenever the discount factor is high enough – so the firm cares sufficiently about future profit – the firm will find it profitable to inflate its costs.[110]

These dynamic price regulation mechanisms affect both the pattern of relative prices and the average price level. The tariff-basket adjustment mechanism reflected in constraint (74) performs well on the first dimension. Starting from some initial price vector, consumer surplus rises monotonically over time and converges to a desirable Ramsey-like pattern of relative prices. However, this mechanism may not control adequately the average price level, and the firm may enjoy positive rent indefinitely. The lagged expenditure policy attempts to deliver a desirable equilibrium pattern of relative prices and to eliminate rent over time. However, it is essentially a form of cost-plus (or rate-of-return) regulation, albeit one that gives the firm flexibility over the pattern of its relative prices. When the firm's cost function is exogenous, the scheme works reasonably well. However, when the firm can affect its production costs, the scheme can provide poor incentives to control costs, and so can induce high average prices.

[110] Sappington (1980) shows that, because of the pure waste it can induce, the lagged expenditure policy may cause welfare to fall below the level that would arise in the absence of any regulation. Hagerman (1990) shows that incentives for pure waste can be eliminated if the policy is modified to allow the firm to make discretionary transfer payments to the regulator. These transfer payments provide a less costly way for the firm to relax the constraint that the lagged expenditure policy imposes on prices.

3.2.2. Non-Bayesian price adjustment mechanisms: transfers

Although Section 3 focuses on policies in which the regulator has no authority to make transfer payments to the regulated firm, we briefly discuss here some non-Bayesian policies that do permit transfers. When transfers are employed, the relevant benchmark entails marginal-cost prices, rather than the Ramsey (and Ramsey-like) prices that were the focus of Section 3.2.1. There are at least three non-Bayesian policies that can eventually implement marginal-cost pricing under certain conditions. The speed of convergence and the distribution of surplus varies considerably under these three mechanisms.

The first such policy was discussed in Section 2.3.1. If the regulator is perfectly informed about consumer demand for the regulated product, he can induce marginal-cost pricing immediately by offering the firm a transfer payment, T, equal to the level of consumer surplus, $v(p)$, generated by the price, p, the firm sets for its product. (For simplicity, assume the firm produces a single product.) By awarding to the firm the entire surplus generated by its actions, this policy induces the firm to maximize total surplus by setting the price equal to its marginal cost.[111] An obvious drawback to this policy is the highly asymmetric distribution of surplus it implements. To recoup some surplus for consumers without distorting the firm's incentive to establish an efficient price, the regulator might subtract a fixed amount (k) from the transfer payment to the firm (so $T = v(p) - k$). Of course, determining an appropriate value for k can be problematic. If k is too small, the firm will continue to enjoy substantial rent. If k is too large, the firm will refuse to operate, and thereby eliminate all surplus.[112]

In a dynamic context, it is possible to return surplus to consumers over time and still maintain marginal-cost pricing. One way to do so is with the following policy.[113] In each period t the regulated firm is permitted to set any price p^t for its product. The regulator pays the firm a transfer each period equal to the difference between the incremental (not the total) consumer surplus derived from its pricing decisions and the firm's profit in the preceding period. Formally, this transfer in period t, T^t, is defined by

$$T^t = \left[v\left(p^t\right) - v\left(p^{t-1}\right)\right] - \Pi^{t-1}, \tag{79}$$

where $v(\cdot)$ is the (known) consumer surplus function and Π^{t-1} is the firm's actual profit in the previous period $t - 1$, which is observed by the regulator. In addition to this transfer, the firm is permitted to keep its profit Π^t in each period.

To illustrate the workings of this policy, termed the *incremental surplus subsidy* policy, suppose there is an exogenous profit function $\pi(p^t)$, the precise form of which is not known to the regulator. (Only actual profit $\Pi^t = \pi(p^t)$ is observed.) In period t,

[111] Loeb and Magat (1979) analyze this regulatory policy.

[112] Section 4.5 discusses how the choice of k may be more straightforward when the regulated firm supplies many products.

[113] This policy is due to Sappington and Sibley (1988).

the firm's pricing decision affects its profit $\pi(p^t)$ in that period, its transfer payment T^t in that period, and its transfer payment in the subsequent period T^{t+1}. In sum, from the perspective of period t, the firm's discounted profit due to its period-t price decision is

$$\pi\left(p^t\right) + v\left(p^t\right) - \delta\left[\pi\left(p^t\right) + v\left(p^t\right)\right] = (1-\delta)\left[\pi\left(p^t\right) + v\left(p^t\right)\right],$$

where δ denotes the firm's discount factor. Therefore, the firm will choose its price to maximize $[\pi(p^t) + v(p^t)]$, which entails marginal-cost pricing in all periods $t = 1, 2, \ldots$. Moreover, from period 2 onward the firm makes zero rent in each period. (Its price is constant, and the firm's stationary operating profit Π is extracted by the transfer in each period.)

Notice also that, in contrast to the lagged expenditure policy, the incremental surplus subsidy policy does not provide incentives for the firm to distort its observed profit. To see why, suppose that when it takes some (unobserved) action e, the firm's realized profit function is $\pi(p, e)$. (For instance, e could take the form of "pure waste", so that $\Pi = \pi(p) - e$ for some "true" profit function π.) Then in period t the firm will choose p^t and e^t in order to maximize $(1-\delta)[\pi(p^t, e^t) + v(p^t)]$, and so the socially efficient level of e^t will be chosen.

In sum[114]:

PROPOSITION 11. *In a stationary environment the incremental surplus subsidy policy ensures*: (i) *marginal-cost pricing from the first period onwards*; (ii) *the absence of pure waste*; *and* (iii) *zero rent from the second period onwards.*

Despite its potential merit in returning surplus to consumers, the policy has at least four drawbacks. First, it can impose financial hardship on the firm if its costs rise over time.[115] Second, the large subsidy payments that the mechanism initially requires are socially costly when the regulator prefers consumer surplus to rent.[116] Third, the regulator must know consumer demand to implement the policy. Finally, although it avoids pure waste, the policy does not preclude "abuse". Abuse is defined as expenditures in excess of minimal feasible costs that provide direct benefit to the firm's managers or employees. Abuse includes perquisites for the firm's managers and the lower managerial effort required to produce at inefficiently high cost, for example.[117]

[114] Schwermer (1994) and Lee (1997b) provide extensions of the incremental surplus subsidy policy to settings with Cournot and Stackelberg competition. Sibley (1989) modifies the scheme to allow the firm to have private information about consumer demand.

[115] See Stefos (1990) and Sappington and Sibley (1990).

[116] Lyon (1996) presents simulations which suggest that once subsidy costs are accounted for, the lagged expenditure policy [modified as Hagerman (1990) suggests to eliminate incentives for pure waste] often generates higher levels of welfare than the incremental surplus subsidy policy.

[117] Sappington and Sibley (1993) show that the mechanism induces the firm's owners to undertake efficient precautions against abuse by subordinates in the firm. However, abuse by the owners themselves can be problematic under the mechanism.

To understand why the regulated firm may undertake abuse under the incremental surplus subsidy policy, consider a case where the regulator can observe some, but not all, components of the firm's costs. Specifically, suppose unit cost c is observed, while the fixed cost $F(c)$, which represents the managerial effort associated with producing with unit cost c, is not observed. Further suppose the transfer payment in period t is

$$T^t = \left[v\left(p^t\right) - v\left(p^{t-1}\right)\right] - Q\left(p^{t-1}\right)\left(p^{t-1} - c^{t-1}\right). \tag{80}$$

In this setting, the firm will choose p^t and c^t to maximize $(1-\delta)[Q(p^t)(p^t - c^t) + v(p^t)] - F(c^t)$. Price p^t will be set equal to realized cost c^t, but c^t will be set at an inefficiently high level.[118] This is because the firm does not retain the full benefit of a cost reduction, since any profit generated in one period is fully usurped in the next period. (Notice from Equation (80) that if, by incurring a high fixed cost, the firm achieves a low marginal cost c^t in one period, it will receive a lower transfer T^{t+1} in the subsequent period. Consequently, the firm will not appropriate the full benefits of its unobserved cost-reducing activity.)

In more realistic settings where realized consumer demand is observed but the regulator does not know the functional form of the demand curve for the firm's product, the exact incremental surplus element of the transfer (79), $[v(p^t) - v(p^{t-1})]$, might be replaced by the linear approximation $q^{t-1}[p^{t-1} - p^t]$, where q^{t-1} denotes realized consumer demand in period $t-1$.[119] Under this policy, the transfer payment to the firm in period t would be

$$T^t = q^{t-1}\left[p^{t-1} - p^t\right] - \Pi^{t-1}. \tag{81}$$

This policy eventually ensures outcomes similar to those induced by the incremental surplus subsidy policy. Thus, if demand and cost functions do not change over time, this mechanism ultimately achieves the outcome a welfare-maximizing regulator would implement if he shared the firm's private knowledge of its environment. However, the convergence of price to marginal cost and the convergence of rent to zero take place only gradually.[120]

3.2.3. *Frequency of regulatory review*

Even when regulatory policies do not explicitly link prices to realized costs, such linkage can be effected when the policies are updated.[121] To illustrate, suppose the authorized rate at which prices can rise (i.e., the X factor) in a price cap regulation regime

[118] This distortion parallels the optimal distortion induced in the Laffont–Tirole setting of Proposition 3 above.

[119] Finsinger and Vogelsang (1981, 1982) proposed this policy (before the incremental surplus subsidy policy was proposed). As originally proposed, this policy was designed to motivate public enterprises. The ensuing discussion adapts the original policy to apply to a profit-maximizing regulated firm.

[120] Finsinger and Vogelsang (1981, 1982) prove this convergence result. Vogelsang (1988) proves that pure waste will not occur.

[121] Explicit linkage of prices to costs is discussed in Section 3.3.

is updated periodically to eliminate the firm's expected future profit. Also suppose expectations about future profit are based in part upon the firm's current realized revenues and costs.[122] Even though a regulatory regime of this sort permits the firm to retain all the profit it generates in any given year, the firm recognizes that larger present profits – generated by efficiency gains, for instance – may result in smaller future earnings. Consequently, implicit intertemporal profit sharing of this sort can limit the firm's incentive to reduce its operating costs and expand its revenues, just as explicit profit-sharing requirements can.

The diminution in incentives will be more pronounced the more frequently the regulatory regime is revised to eliminate expected rent. On the other hand, an infrequent revision of the regime could allow prices to diverge from costs for long periods, and thereby reduce allocative efficiency. The optimal choice of "regulatory lag" trades off these two opposing effects.[123] The following extreme settings provide some intuition for the key determinants of the optimal frequency of regulatory review:

- If the firm cannot affect its realized costs, frequent regulatory reviews are optimal. Because there is no need to provide incentives for cost reduction in this case, the only concern is to achieve allocative efficiency. When costs vary over time, this goal is best accomplished through frequent reviews that set prices to match realized costs.
- If consumer demand is inelastic, so there is little deadweight welfare loss when prices depart from costs, reviews should be infrequent. If prices are permitted to diverge from realized costs for long periods of time, the firm will have strong incentives to reduce costs, since the firm keeps most of the extra surplus it generates. And when there is little efficiency gain from ensuring that prices track costs closely, it is optimal to implement long lags between regulatory reviews.

Clearly, any realistic case will fall between these two extremes, and the optimal period between reviews in a price cap regime will depend upon the specifics of the regulatory environment. A key element of this environment is the regulator's ability to implement credible long-term contracts. If formal regulatory reviews are carried out only infrequently, the firm's realized profits can become too large or too small for the regulator

[122] When implemented in this manner, price cap regulation operates much like rate-of-return regulation with a specified regulatory lag. Baumol and Klevorick (1970) and Bailey and Coleman (1971), among others, analyze the effects of regulatory lag on incentives for cost reduction under rate-of-return regulation. Pint (1992) examines the effects of regulatory lag under price cap regulation and demonstrates the importance of basing projections of future costs on realized costs throughout the price cap regime rather than in a single test year. When a test year is employed, the regulated firm can limit its cost-reducing effort in the test year and shift costs to that year in order to relax future regulatory constraints.

[123] This discussion is based on Armstrong, Rees and Vickers (1995). Notice that the choice of an infrequent regulatory review may enable the regulator to commit to remaining partially ignorant of the firm's costs. This ignorance allows the regulator to promise credibly not to link prices too closely to costs, even when he cannot commit to future pricing policies. (Recall the discussion of an analogous strategy in Section 2.5.) Isaac (1991) points out that rate shock (substantial, rapid price changes) may occur if prices are revised to reflect realized operating costs only infrequently.

to refrain from intervening to re-set prices before the scheduled review. In such cases, long regulatory lags are not credible.

The lagged expenditure policy discussed in Section 3.2.1 can be viewed as a regulatory regime with frequent regulatory reviews. With this policy, the firm's prices in each period are required to fall to a level that reflects realized expenditures in the previous period. As noted, this mechanism can provide poor incentives to control costs, even though it serves to implement desirable prices given the realized costs. More generally, the frequency of regulatory review is essentially a choice about how responsive prices will be to realized costs. This issue is explored further in Section 3.3.

3.2.4. *Choice of 'X' in price cap regulation*

Recall from the discussion in Section 3.2.1 that it may be desirable to require the (inflation-adjusted) prices charged by a regulated firm to decline at a specified rate, X. In practice, it can be difficult to determine the most appropriate value of this "X factor". To provide some insight regarding the appropriate choice of an X factor, consider a setting where (in contrast to the preceding discussion of dynamic regulatory policies) the firm invests in durable capacity over time. To simplify the analysis, suppose there is no asymmetric information and the regulated firm produces a single product.[124]

Further suppose that investment, production and consumption all take place in periods $t = 0, 1, \ldots$. Let p_t denote the price for the firm's product in period t. Suppose that consumer surplus and the demand function for the firm's product in period t are, respectively, $v_t(p_t)$ and $Q_t(p_t)$. For simplicity, demand in each period is assumed to depend only on the price set in that period. Over time, the firm invests in the capacity required to deliver its product. For simplicity, one unit of capacity is assumed to be needed to provide one unit of service. Capacity at time t is denoted K_t. Capacity depreciates at the proportional rate d in each period. The cost of installing a unit of capacity in period t is β_t, so there are constant returns to scale in installing capacity. Let I_t be the investment (in money terms) undertaken in period t, so the amount of new capacity installed in period t (in physical units) is I_t/β_t. Therefore, capacity evolves according to the dynamic relation

$$K_{t+1} = (1-d)K_t + \frac{I_{t+1}}{\beta_{t+1}}. \tag{82}$$

All investment can be used as soon as it is installed.

What is the marginal cost of providing an extra unit of service in period t in this setting? Suppose the investment plan is $K_t, K_{t+1}, \ldots, I_t, I_{t+1}, \ldots$, satisfying expression (82). Then if K_t is increased by 1, all subsequent values for K and I are unchanged

[124] The analysis in this section is based on Laffont and Tirole (2000, Section 4.4.1.3). See Kwoka (1991, 1993), Armstrong, Cowan and Vickers (1994, Section 6.3), and Bernstein and Sappington (1999) for further discussions.

if next period's investment I_{t+1} is reduced so as to the keep the right-hand side of expression (82) constant, i.e., if I_{t+1} is reduced by $(1-d)\beta_{t+1}$.[125] If the interest rate is r, so that the firm's discount factor is $\delta = \frac{1}{1+r}$, then the net cost of this modification to the investment plan is

$$C_t = \beta_t - \frac{1-d}{1+r}\beta_{t+1}. \tag{83}$$

Expression (83) specifies the marginal cost of obtaining a unit of capacity for use in period t. If technical progress causes the unit cost of new capacity to fall at the exogenous rate γ every period, then $\beta_{t+1} = (1-\gamma)\beta_t$. With technical progress at the rate γ, formula (83) becomes[126]

$$C_t = \beta_t\left(1 - \frac{(1-d)(1-\gamma)}{1+r}\right). \tag{84}$$

Clearly, this marginal cost of capacity falls (with β_t) at the rate γ.

Suppose it costs the firm an amount c_t to convert a unit of capacity into a unit of the final product. Then the total marginal cost for supplying the final product is $C_t + c_t$, and so the optimal price in period t is $p_t = C_t + c_t$, where C_t is defined in expression (84). Thus, in this setting with constant returns to scale, welfare is maximized if, in each period, price is set equal to the correctly calculated marginal cost of expanding available capacity for one period, C_t, plus the operating cost c_t. If both the cost of capacity β_t and the operating cost c_t fall at the same exogenous rate γ, then this optimal price should also fall at this rate γ, i.e., 'X' should be equal to the exogenous rate of technical progress.

Of course, the cost structure of the regulated firm and the (potential) rate of technical progress are unlikely to be common knowledge in practice.[127] To secure a modest X, the regulated firm may claim that its potential for cost reduction and the rate of technical progress are modest. In contrast, consumer advocates are likely to argue that the firm is capable of achieving pronounced productivity gains. In practice, a regulator is forced to weigh the available evidence, however limited it might be, and make his best judgment about a reasonable value for the X factor.

3.3. The responsiveness of prices to costs

The discussion in Section 3.2 emphasized the importance of the extent to which regulated prices are (implicitly or explicitly) linked to costs. The present section considers

[125] Assume that demand conditions are such that investment in each period is strictly positive, which ensures that this modification is feasible.

[126] If the parameters d, r and γ are reasonably small, this formula is approximated by $C_t \approx \beta_t(r+\gamma+d)$. This is a familiar equation in continuous-time investment models.

[127] In addition, in practice there is considerable uncertainty (for both the regulator and firm) about how consumer demand and technology develop over time. See Dobbs (2004), for instance, for an account of how the principles of dynamic price regulation are altered in the presence of uncertainty. Guthrie (2006) provides a survey of the literature that examines the effects of regulatory policy on investment.

this linkage in more detail, and explores the trade-offs involved in varying the extent to which prices reflect realized costs. The focus in this section is on the trade-off between allocative efficiency and providing incentives for the firm to control its costs. The discussion employs the moral hazard framework of Section 2.6.

Recall from Section 2.6.1 that when transfer payments between the regulator and the firm are possible and the firm is risk neutral, consumers are best served by affording the firm the entire social gains that its unobserved activities secure. The reason is that incentive issues are resolved fully when the firm is the residual claimant for the surplus it generates, and the firm can be required to compensate consumers in advance for the right to retain the incremental surplus it generates, which resolves the distributional issue. This conclusion suggests that high-powered incentive schemes like price cap regulation are better suited for resolving moral hazard problems than are low-powered policies like rate-of-return regulation, at least when risk aversion, limited liability, and asymmetric knowledge of the firm's production technology are not serious concerns. It is useful to examine how this conclusion is modified when transfer payments from the regulator to the firm are not possible.

For simplicity, consider the moral hazard setting where marginal cost can be either high or low, so that the firm's profit function in state i is $\pi_i(p) \equiv Q(p)(p - c_i)$. The equilibrium probability of achieving a low-cost outcome, $\hat{\phi}(\Delta^U)$, is given by expression (55) above. (Recall $\Delta^U = U_L - U_H$ is the difference between the firm's utility in the low state and the high state.) Suppose further that the demand function is iso-elastic, with constant elasticity equal to η.[128] Suppose that transfer payments are prohibitively costly, so the regulator can only dictate the unit price the firm will be allowed to charge given its realized costs.[129]

In this setting, prices are required to perform two tasks. First, they must provide the firm with incentives to reduce costs. To provide such incentives, the firm's profit must be higher when its costs are lower. Second, prices must not depart too far from realized cost in order to promote allocative efficiency. Clearly, ideal incentives and allocative efficiency cannot be achieved simultaneously, and a compromise is required.

It follows from Equations (61) and (62) that the full-information prices in this setting (i.e., the prices the regulator would allow the firm to choose if the regulator could directly control the firm's cost-reducing effort) are

$$\frac{p_L - c_L}{p_L} = \frac{p_H - c_H}{p_H} = \left[\frac{\lambda}{1+\lambda}\right]\frac{1}{\eta}, \tag{85}$$

[128] If demand is inelastic so $\eta \leqslant 1$, suppose that demand goes to zero when price reaches some high "choke price" in order to make consumer surplus well defined. This choke price is assumed to be higher than any of the prices identified in the following discussion.

[129] Implicitly, we rule out both transfer payments from taxpayers to the firm and two-part tariffs. The following discussion is closely related to Schmalensee (1989). His model differs in that a continuum of cost states are possible and he restricts attention to linear incentive schemes. (This restriction is inconsequential when there are just two possible outcomes.) He also models the regulator as being uncertain about the cost of effort function for the firm. See Gasmi, Ivaldi and Laffont (1994) for further analysis of a similar model.

where, recall, λ is the Lagrange multiplier associated with the firm's participation constraint. Thus, the Lerner index $(p_i - c_i)/p_i$ is equal for the two cost realizations, in accordance with standard Ramsey principles. (See expression (2) above.) At this full-information outcome, prices vary proportionally with realized costs. The resulting relationship between profit and the cost realization depends on the demand elasticity: with equal mark-ups, the firm's profit $\pi_i(p_i)$ is higher (respectively lower) when costs are low if $\eta > 1$ (respectively if $\eta < 1$). Thus, when demand is inelastic, the firm makes less profit when its costs are low under the full-information policy. Of course, such a policy provides no incentive for the firm to achieve a low cost.

Turning to the second-best problem where the regulator cannot directly control the firm's cost-reducing effort, expression (63) in the present setting becomes

$$\frac{p_L - c_L}{p_L} = \left[\frac{\hat{\lambda}\hat{\phi} + \Delta^V \hat{\phi}'}{(1+\hat{\lambda})\hat{\phi} + \Delta^V \hat{\phi}'}\right]\frac{1}{\eta},$$
$$\frac{p_H - c_H}{p_H} = \left[\frac{\hat{\lambda}(1-\hat{\phi}) - \Delta^V \hat{\phi}'}{(1+\hat{\lambda})(1-\hat{\phi}) - \Delta^V \hat{\phi}'}\right]\frac{1}{\eta}. \tag{86}$$

Here, $\Delta^V = v(p_L) - v(p_H)$ is the difference in consumer surplus in the two states at the optimum.

As in Section 3.2.3, it is useful to consider two extreme cases:

- If the success probability ϕ is exogenous, there is no need to motivate the firm to achieve lower production costs. (In this case, $\hat{\phi}'(\cdot) = 0$ and expressions (86) reduce to the standard full-information Ramsey formulas (85).) Thus, pure cost-plus regulation is optimal in this setting.
- If demand is perfectly inelastic, there is no welfare loss when price diverges from cost. Consequently, in this setting, it is optimal to provide the maximum incentive for cost reduction. This is accomplished by setting a price that does not vary with realized costs (so $p_L = p_H$). In this case, it is optimal to implement pure price cap regulation, and the full-information outcome is achieved again.[130]

In less extreme cases, departures from the full-information policy are optimal. Expressions (86) imply that[131]

[130] This is essentially an instance of the analysis of optimal regulation with a risk-neutral firm when transfers are used, discussed in Section 2.6.1. When demand is perfectly inelastic, there is no difference between the use of prices and transfers, and a prohibition on the regulator's use of transfers is not restrictive.

[131] If $p_L \leqslant p_H$ then $\Delta^V \geqslant 0$. In this case, expressions (86) imply that

$$\frac{p_L - c_L}{p_L} \geqslant \left[\frac{\hat{\lambda}}{1+\hat{\lambda}}\right]\frac{1}{\eta} \geqslant \frac{p_H - c_H}{p_H},$$

so the Lerner index is higher for the low-cost firm. On the other hand, if $p_L > p_H$, expression (86) implies

$$\frac{p_L - c_L}{p_L} < \frac{p_H - c_H}{p_H},$$

which is clearly inconsistent with $p_L > p_H$. Therefore, the only possible configuration consistent with (86) is as given in expression (87).

$$\frac{p_L - c_L}{p_L} \geqslant \frac{p_H - c_H}{p_H}, \qquad p_L \leqslant p_H \tag{87}$$

and so the Lerner index is higher in the low-cost state than in the high-cost state in order to provide an incentive for cost reduction. In particular, regulated prices do not fully reflect costs, although the regulated price declines as realized cost declines. To provide strong incentives to reduce cost, it can even be optimal to set price below cost when the high cost is realized.

In summary, when the regulator cannot make transfer payments to the firm, prices are required to pursue both allocative efficiency and productive efficiency. The inevitable compromise that ensues results in prices that are higher when realized costs are low than they would be in a full-information world. The higher prices motivate the firm to secure low costs.[132]

This discussion has so far been confined to a pure moral hazard framework, in which the firm has no private information at the time regulatory policy is formulated. A richer framework would allow the firm to possess private information about its ability to obtain low costs, for instance. In such cases, the regulator might offer the firm a choice among regulatory plans. To understand the nature of this choice, consider a benchmark case of pure adverse selection, where the firm knows its probability of achieving a low cost realization, but has no ability to affect this probability. (The firm does not know the actual cost realization at the time the regulatory policy is determined, however.) Suppose the firm is either type L or type H. The type-L firm has a higher probability of achieving low cost than the type-H firm. In this setting where transfer payments are not feasible, the regulator will offer the firm a choice between two pairs of prices, (p_L^L, p_H^L) and (p_L^H, p_H^H), where p_i^j denotes the regulated price when the firm claims to be type j and realized (and observed) cost is c_i. These prices will be designed to ensure the type-H firm enjoys no expected rent while the type-L firm is indifferent between the two pairs of prices. To make the (p_L^H, p_H^H) option less attractive to the type-L firm and thereby to reduce the type-L firm's rent, p_L^H will be reduced below the Ramsey levels identified in expression (85). Because the type-L firm is more likely to achieve cost c_L than the type-H firm, the reduction in p_L^H (and corresponding increase in p_H^H) is differentially unattractive to the type-L firm. The regulator implements no corresponding distortions from Ramsey prices in the (p_L^L, p_H^L) option, because such distortions would not reduce the rent of the type-H firm, which strictly prefers the (p_L^H, p_H^H) option to the (p_L^L, p_H^L) option.[133]

Richer models could incorporate both moral hazard and adverse selection. For instance, the firm might have private information about its ability to reduce costs. As the analysis in Section 2 suggests, a carefully structured choice among regulatory plans can

[132] This analysis is closely related to the analysis in Section 2.6.2, where transfer payments are possible and the firm is risk averse. In both cases, a concave relationship between consumer surplus and the firm's utility makes the optimal regulatory policy less high powered than the full-information policy.
[133] See Armstrong and Vickers (2000, Section IV) for a model along these lines.

limit the regulated firm's incentive to understate its potential to achieve productivity gains. To illustrate, the firm might be afforded the choice between: (1) a pure price cap plan; and (2) a plan where prices reflect realized costs to some extent. When the parameters of these plans are chosen appropriately, the firm can be induced to select: (1) the pure price cap plan when it knows that it has pronounced ability to reduce its operating costs; and (2) the earnings sharing plan when it knows that its ability to reduce operating costs is more limited.[134] The more capable firm is willing to guarantee lower prices to consumers in return for the expanded opportunity to retain more of the relatively high earnings it knows it can generate. The less capable firm is willing to share its (relatively modest) earnings with its customers when doing so allows it to guarantee more modest price reductions.

3.4. Regulatory discretion

The final key element of the design of regulatory policy that will be considered here is the degree of policy discretion afforded the regulator. When the regulator has extensive, ongoing experience in the industry, he will often be well informed about relevant industry conditions, in which case it can be advantageous to afford him considerable latitude in policy design. However, a regulator might employ this latitude inappropriately. In particular, the regulator might behave opportunistically over time, maximizing welfare ex post in such a way as to distort the ex ante incentives of the firm. Alternatively, the regulator might succumb to industry pressure to act in a non-benevolent manner. These two dangers are discussed in turn.

3.4.1. Policy credibility

Section 2.5.3 explained how a regulator's inability to commit to future policy can harm the regulatory process. The key problem in Section 2.5.3 was that the regulator could not refrain from using information revealed early in the process to maximize subsequent welfare. Another fundamental problem arises in the presence of sunk investments.[135] Once the firm has made irreversible investments, a regulator with limited commitment powers may choose not to compensate the firm for those investments, in an attempt to

[134] See Laffont and Tirole (1986) and Lewis and Sappington (1989d) for formal analyses of this issue, and Sappington and Weisman (1996a, pp. 155–165) for further discussion. Rogerson (2003) provides conditions under which a particularly simple regulatory policy secures a large fraction (at least three-quarters in a parameterized example) of the surplus secured by the optimal regulatory policy. The simple regulatory policy consists of only two options: a pure price cap plan and a plan under which the firm's realized costs are fully reimbursed. Bower (1993), Gasmi, Laffont and Sharkey (1999), McAfee (2002), and Chu and Sappington (2007) also analyze settings in which a limited number of options and/or simple contracts secure much of the surplus that a richer set of options can secure.

[135] See Williamson (1975) for a pioneering treatment of the problem, and Newbery (1999, ch. 2) for a detailed discussion of the problem of regulatory commitment. Tirole (1986b) considers both the information and investment aspects of the commitment problem.

deliver the maximum future benefits to consumers. This expropriation might take the form of low mandated future prices. Alternatively, the expropriation might arise in the form of permitting entry into the industry.[136] When it anticipates expropriation of some form, the firm will typically undertake too little investment.[137]

One natural way to overcome the temptation for a regulator to behave opportunistically is to limit the regulator's policy discretion. This might be done, for instance, by imposing a legal requirement that the firm earn a specified rate of return on its assets.[138] The promise of a fair return on investment can provide relatively strong incentives for infrastructure investment in a dynamic setting where the regulator has weak commitment powers. However, a blanket commitment to deliver a specified return on assets can reduce significantly the firm's incentives to control its costs, in part because the commitment rewards inefficient or unnecessary projects in the same way it rewards efficient projects. To limit this problem, the naive rate-of-return commitment could be modified to consider whether the assets are ultimately "used and useful". There are two problems with such a policy, though. First, an investment might ultimately prove to be unnecessary even though it was originally desirable. The merits of a given investment often are difficult to predict precisely, in practice. Second, if the regulator has some discretion in defining which sunk investments are included in the asset base, the problem of limited regulatory commitment resurfaces.[139]

[136] Price cap regulation can encourage the regulator to expropriate the incumbent firm by introducing competition. Recall that under price cap regulation, prices are not linked explicitly to the earnings of the regulated firm. In particular, the regulator is under no obligation to raise prices in the regulated industry if the firm's profit declines. This fact may encourage the regulator to facilitate entry into the industry in order to secure even lower prices for consumers. The regulator may be more reluctant to encourage entry under rate-of-return regulation because he might then be obliged to raise industry prices in order to mitigate any major impact of entry on the profits of the incumbent firm – see Weisman (1994). Lehman and Weisman (2000) provide some empirical support for this effect. Kim (1997) analyzes a model in which a welfare-maximizing regulator decides whether entry should be permitted once the incumbent has made investment decisions. Biglaiser and Ma (1999) find that entry into a regulated industry where the regulator's commitment powers are limited can either enhance or diminish incentives for cost-reducing investment by the incumbent firm. The direction of the effect depends upon how investment affects the distribution of the firm's operating costs.

[137] Spiegel (1994) and Spiegel and Spulber (1994, 1997) demonstrate how the regulated firm may alter its capital structure in order to induce a regulator with limited commitment power to authorize a higher regulated price. Specifically, the firm may choose a high debt-equity ratio in order to make bankruptcy – which involves extra costs that the regulator takes into account when determining future pricing policy – more likely for a given price policy. To avoid the costs of bankruptcy, the regulator implements a more generous pricing policy than he otherwise would.

[138] See Greenwald (1984). Levy and Spiller (1994) and Sidak and Spulber (1997) examine the legal framework governing a regulator's ability to expropriate a firm's sunk investments.

[139] See Kolbe and Tye (1991), Lyon (1991, 1992), Gilbert and Newbery (1994), and Encinosa and Sappington (1995) for analyses of regulatory cost disallowances and "prudence reviews". Sappington and Weisman (1996b) examine how the discretion of the regulator to disallow certain investments affects the firm's investment decisions.

Although limited regulatory commitment can discourage investment, it need not always do so.[140] When the regulator and firm interact repeatedly, mutual threats by the firm and regulator to "punish" one another can sustain desirable investment and compensation levels.[141] To illustrate, in a model where investments last forever – which is where the danger of expropriation is especially great – desired investment levels can ultimately be achieved if the firm gradually builds up its asset base. Here, if the regulator reneges on his implicit promise to deliver a reasonable return on capital, the firm can punish the regulator by refusing to continue its capital expansion program.[142]

Institutional design also can enhance regulatory commitment powers. For instance, a government might intentionally employ a regulator who values industry profit (relative to consumer surplus) more highly than the government. Such a regulator will be relatively unlikely to expropriate industry investment, and so valued investment is relatively likely to be undertaken.[143] The division of regulatory responsibility among multiple regulators, each with a different objective, also may help to enhance regulatory commitment powers. Absent commitment problems, the conflicting objectives of multiple regulators can complicate policy design and implementation.[144] However, the conflicting objectives and dispersed powers can limit the incentive and ability of any single regulator to renege on a promise he has made, and thereby enhance incentives for the firm to undertake valued investment.[145]

[140] Besanko and Spulber (1992) demonstrate that a regulated firm may undertake excessive investment to induce an opportunistic regulator to set a higher price for the firm's product. In the model, the regulator is uncertain about the relationship between the firm's observable capital stock and its unobservable unit operating cost. In equilibrium, higher levels of capital lead the regulator to increase his estimate of the firm's unit cost of operation. Consequently, the firm undertakes more than the cost-minimizing level of capital investment to induce the regulator to revise upward his estimate of the firm's operating cost, and to set a correspondingly higher price for the firm's product.

[141] Of course, this is just an instance of the general theory of dynamic and repeated games. See Fudenberg and Tirole (1991, ch. 5) for an overview. Gilbert and Newbery (1994) and Newbery (1999, ch. 2) compare the abilities of three kinds of regulatory contracts to induce desirable investment in the presence of limited regulatory commitment: (i) naive rate-of-return regulation, (ii) rate-of-return regulation with a "used and useful" requirement, and (iii) price-cap regulation. Consumer demand is uncertain in their model, and so capacity investment that is desirable ex ante may not be required ex post. The authors show that regime (ii) can sustain the desirable rate of investment for a larger range of parameter values than either regime (i) or regime (iii). Lewis and Sappington (1990, 1991b) assess the merits of alternative regulatory charters.

[142] See Salant and Woroch (1992) for a formal analysis of this issue, and see Levine, Stern and Trillas (2005) for a model based on regulatory reputation. Lewis and Yildirim (2002) show that learning-by-doing considerations can limit incentives for regulatory expropriation. When higher present output reduces future operating costs, a regulator may persistently induce greater output from, and thereby provide more rent to, the regulated firm than in settings where present output levels do not affect future costs.

[143] Again, see Levine, Stern and Trillas (2005).

[144] See Baron (1985), for example.

[145] To illustrate this possibility, consider the possible benefits of private versus public ownership of an enterprise as discussed in Laffont and Tirole (1991c) and Laffont and Tirole (1993b, ch. 17). Under public ownership, the government tends to use assets for social goals instead of for profit, and so a commitment problem may arise. With (regulated) private ownership, however, the firm's manager has two bodies con-

A regulator's incentive to expropriate sunk investments also may be tempered by increasing the political cost of such expropriation. This political cost might be increased, for example, by privatizing and promoting widespread ownership of a regulated firm. (The widespread ownership might be accomplished by setting low share prices, restricting the number of shares an individual can own, and providing long-term incentives not to sell the shares.) Such widespread ownership can help to ensure that a large fraction of the population will be harmed financially if the regulator expropriates the firm. Such widespread harm can increase the political cost of expropriation, and thereby limit its attraction to the regulator.[146]

The foregoing discussion presumes regulatory commitment is desirable. The ability to commit generally will be desirable when the regulator naturally pursues long-term social objectives. However, limited commitment may be preferable when the regulator is susceptible to capture or is myopic. To see why, consider a simple dynamic setting in which a regulator is susceptible to capture in each period with some exogenous probability. Suppose the government can decide whether to allow the regulator to write long-term contracts with the firm in this setting, i.e., whether the regulator can commit to future policy. Endowing the regulator with such commitment power involves a trade-off: commitment can enable the regulator to promise credibly not to expropriate the firm's sunk investments (and thereby encourage such investments), but it also allows a captured regulator to commit to policies that harm consumers. Whether commitment is desirable in such a setting can depend in complicated ways on model parameters. For instance, commitment is desirable if the probability of capture is small (as one would expect). However, commitment can also be desirable if capture is very likely.[147,148] Similar considerations arise when the regulator may act myopically. For instance, a regulator might have a relatively short term of office and maximize the welfare only over this term, ignoring the effects of his actions after his term has ended. In this case, the

trolling him: the regulator (who is interested in maximizing future welfare) and shareholders (who seek to maximize profit). These two bodies simultaneously offer the manager an incentive scheme, rewarding him on the basis of his performance. The equilibrium of this game between shareholders and the regulator determines the manager's actions. Joint control can produce a higher level of investment than is secured under unilateral control by government, and so can mitigate the commitment problem that exists under public ownership. See Martimort (1999) for further analysis of how multiple regulators can lessen a regulator's temptation to renegotiate contracts over time.

146 See Vickers (1991), Schmidt (2000) and Biais and Perotti (2002) for formal analyses of this issue.

147 Laffont and Tirole (1993b, ch. 16) analyze this model. The comparative statics with respect to the probability of capture are ambiguous because there are two conflicting effects. To see why, suppose, for simplicity, there are two periods and regulators are short-lived. If capture is unlikely, then it generally is desirable to allow the initial regulator to write long-term contracts in order to induce efficient long-term investment by the firm. However, when capture is unlikely, it is also likely that the second-period regulator will be honest, and will correct any bad policy made in the first period (in the unlikely event that the initial regulator was corrupt). This latter effect tends to make short-term contracts more desirable.

148 Faure-Grimaud and Martimort (2003) show that despite the danger of capture by the regulated firm, a government may grant a regulator some long-term independence in order to limit the influence of future governments (with different political interests) on industry policy.

ability to write long-term contracts may be undesirable because it can allow the regulator to pass excessive costs on to future generations.[149]

3.4.2. Regulatory capture

In the model of regulatory capture analyzed in Section 2.4.2, the optimal response to the danger of collusion was (i) to provide the regulator with countervailing incentives to act in the interests of society, and (ii) to reduce the firm's benefit from capturing the regulator. That model proposed what might be termed a "complete contracting" response to the capture problem, as the "constitution" provided the regulator with explicit monetary incentives to behave appropriately. In practice, such detailed contingencies can be difficult to design and implement. Instead, a constitution may only authorize or preclude certain regulatory actions. In such an "incomplete contracting" setting, a constitution might simply prohibit regulators from future employment in the industries they oversee in order to limit regulatory capture.[150] Alternatively, a constitution might preclude transfer payments between the regulator and firm for the same reason.[151] To see why in a specific setting, suppose the regulated firm's fixed cost initially is unknown. If transfer payments from taxpayers to the firm are possible, then marginal-cost pricing is feasible, which enhances allocative efficiency. If transfers are not possible, then average-cost pricing must be pursued.[152] If the regulator is captured, and thus allows an exaggerated report of the firm's fixed costs to be used as the basis for setting tariffs, then: (i) when transfers are used, the large fixed costs are covered by taxpayers and are not reflected in prices, and so go largely unnoticed by consumers; whereas (ii) when average-cost pricing is used, consumers may be acutely aware of any report of high costs by the firm/regulator, since high costs translate into higher prices. If consumers are better organized (or more observant) than taxpayers, then average-cost pricing may result in greater monitoring of the regulator, and hence act as a more effective impediment to capture. In this case, the beneficial effects of a reduced likelihood of capture could outweigh the allocative inefficiencies introduced by the use of average-cost pricing.

[149] See Lewis and Sappington (1990) for an analysis of this issue.

[150] However, Che (1995) shows that the possibility of future employment at a regulated firm can induce regulators to work more diligently during their tenure as regulators. Che also shows that some collusion between the regulator and firm might optimally be tolerated in order to induce the regulator to monitor the firm's activities more closely (in the hopes of securing a profitable side contract with the firm). Also see Salant (1995) for an analysis of how non-contractible investment could be encouraged when the regulator may later be employed by the firm.

[151] This discussion is based on Laffont and Tirole (1990c) and Laffont and Tirole (1993b, ch. 15). For a theory of why transfers should not be permitted that depends on regulatory failures related to commitment problems, see Laffont and Tirole (1993b, pp. 681–682).

[152] There is therefore a restriction to linear pricing in the no-transfer case.

3.5. Other topics

3.5.1. Service quality

To this point, the discussion of practical regulatory policies has abstracted from service quality concerns. In practice, regulators often devote substantial effort to ensuring that consumers receive high-quality regulated services. Before concluding this section, some practical policies that can help to secure appropriate levels of quality for regulated services are discussed briefly.[153]

To understand the basic nature of many practical policies that might be employed to secure appropriate levels of service quality, consider first the levels of service quality that an unregulated monopolist will supply when it can deliver different levels of quality to different consumers. An unregulated monopolist that sells its products to consumers with heterogeneous valuations of quality will tend to deliver less than the welfare-maximizing level of quality to consumers who have relatively low valuations of quality. This under-supply of quality to low-valuation customers enables the monopolist to extract more surplus from high-valuation customers. It does so by making particularly unattractive to high-valuation customers the variant of the firm's product that low-valuation consumers purchase. Faced with a particularly unattractive alternative, high-valuation customers are willing to pay more for a higher-quality variant of the firm's product.[154]

This pattern of quality supply by an unregulated monopolist suggests regulatory policies that might increase welfare. For example, a minimum quality requirement might increase toward its welfare-maximizing level the quality delivered to low-valuation customers. A price ceiling might also preclude the firm from charging high-valuation customers for the entire (incremental) value that they derive from the high-quality variant of the firm's product. Consequently, the firm's incentive to under-supply quality to low-valuation customers may be reduced.[155] And substantial profit taxes can also limit the financial benefits the firm perceives from under-supplying quality to low-valuation customers in order to secure greater profit from serving high-valuation customers.[156]

Price cap regulation alone generally does not provide the ideal incentives for service quality enhancement. Under price cap regulation, the regulated firm bears the full costs

[153] See Sappington (2005b) for a more detailed review of the literature on service quality regulation.

[154] See Mussa and Rosen (1978) for the seminal work in this area.

[155] See Besanko, Donnenfeld and White (1987, 1988) for analyses of these policies. See Ronnen (1991) for an analysis of the merits of minimum quality requirements in a setting where the prices set by competing firms are not regulated. Crampes and Hollander (1995) and Scarpa (1998) provide related analyses.

[156] Kim and Jung (1995) propose a policy that includes lagged profit taxes, and demonstrate that the policy can induce a firm to deliver the welfare maximizing level of service quality to all consumers, provided the firm does not undertake strategic abuse. (Recall from Section 3.2.2 that abuse entails expenditures in excess of minimum feasible costs that provide direct benefit to the firm.) Lee (1997a) proposes a modified policy with lower tax rates that limits incentives for abuse.

of increasing quality, but the price cap constraint prevents the firm from recovering the full value that consumers derive from the increased quality. Therefore, the firm generally will have insufficient incentive to deliver the welfare-maximizing level of service quality. Consequently, price cap regulation plans often incorporate explicit rewards and penalties to ensure the delivery of desired levels of service quality.[157]

When the regulated firm is privately informed about its costs of providing service quality on multiple dimensions, welfare gains can be secured by presenting the firm with a schedule of financial rewards and penalties that reflect the gains and losses that consumers incur as service quality varies on multiple dimensions.[158] In essence, such a schedule, coupled with a policy like price cap regulation that divorces regulated prices from costs, induces the firm to internalize the social benefits and costs associated with variations in the service quality it delivers.[159] Consequently, the schedule can induce the firm to minimize its costs of delivering service quality and to deliver to customers the levels of service quality on multiple dimensions that they value most highly.

3.5.2. Incentives for diversification

Firms that operate in regulated markets often participate in unregulated markets as well. For example, regulated suppliers of basic local telephone service often supply long distance telephone service and/or broadband Internet services at unregulated rates. Additional policy considerations arise when a firm operates, or has the opportunity to operate, simultaneously in both regulated and unregulated markets.

In particular, regulatory policy can affect the incentives of regulated firms to diversify into unregulated markets. To illustrate, suppose a firm operates under a cost-based regulatory policy (like rate-of-return regulation) in which the prices of the firm's regulated services are set to generate revenue that just covers the firm's costs of producing the regulated services. Suppose further that these costs include a portion of the shared (e.g., overhead) costs that arise from the production of both regulated and unregulated services. If the fraction of shared costs that are allocated to regulated operations declines as the firm's output in non-regulated markets increases, the firm typically will produce less than the welfare-maximizing level of unregulated services. This under-supply of

[157] See Laffont and Tirole (2000, p. 88). Spence (1975) and Besanko, Donnenfeld, and White (1987, 1988) note that price cap regulation may diminish the firm's incentive to deliver service quality relative to rate-of-return regulation when the provision of quality is capital intensive. Weisman (2005) points out that penalties for insufficient service quality that are imposed as reductions in the share of realized revenue to which a firm is entitled can reduce a firm's incentive to deliver service quality.

[158] See Berg and Lynch (1992) and Lynch, Buzas and Berg (1994). De Fraja and Iozzi (2004) demonstrate how a regulator that is well informed about consumers' marginal valuations of quality can modify the lagged expenditure policy discussed in Section 3.2.1 to induce a regulated monopoly to set welfare-maximizing prices and quality levels.

[159] Such a policy thereby acts much like the policy proposed by Loeb and Magat (1979), which provides financial rewards to the firm that reflect the level of consumer surplus its performance generates.

unregulated services arises because the cost allocation procedure effectively taxes the firm's output of unregulated services, which reduces their supply.[160]

In contrast, a regulated firm may undertake excessive expansion into unregulated markets if it is able to engage in cost shifting. Cost shifting occurs when the regulator counts as costs incurred in producing regulated services costs that truly arise solely from the production of unregulated services. Under cost-based regulation, cost shifting forces the customers of regulated services to bear some of the costs of the regulated firm's operation in unregulated markets, which explains the excessive expansion of these operations.[161]

Regulated firms that operate in both regulated and unregulated markets also may adopt inefficient production technologies. Technologies that entail particularly high fixed, shared costs and particularly low incremental costs of producing unregulated services can be profitable for a firm that operates under a form of cost-based regulation that attributes most or all shared costs to regulated operations.[162]

Although operations in unregulated markets can harm consumers of regulated services by admitting cost shifting and encouraging inefficient production technologies, diversification into unregulated markets also can benefit regulated customers. The benefits can flow from cost reductions in regulated markets that arise from economies of scope in producing regulated and unregulated services, for example.[163] The opportunity to pursue profit from unregulated operations may also induce a firm to undertake more research and development than it does absent diversification, to the benefit of customers of regulated services.[164]

A regulator also can secure gains for regulated customers by linking the firm's earnings from diversified operations to the welfare of regulated customers. To illustrate, suppose the regulator allows the firm to share the incremental consumer surplus that its diversified operations generates for consumers of the firm's regulated product. (The incremental surplus may arise from price reductions that are facilitated by economies of scope in the production of regulated and unregulated services, for example.) Such a policy, which is feasible when consumer demand for the regulated service is known, can induce the regulated firm to minimize its production costs and to diversify into a competitive unregulated market only when doing so increases aggregate welfare.[165]

A regulator also can secure gains for regulated customers by controlling directly the level of the regulated firm's participation in unregulated markets. To illustrate this fact, consider a variant of Baron and Myerson's (1982) model in which the regulated firm

[160] See Braeutigam and Panzar (1989), Weisman (1993), and Chang and Warren (1997) for formal analyses of this phenomenon.

[161] See Brennan (1990) and Brennan and Palmer (1994).

[162] See Baseman (1981), Brennan (1990), and Crew and Crocker (1991).

[163] Brennan and Palmer's (1994) investigation of the likely benefits and costs of diversification by regulated firms includes an analysis of the potential impact of scope economies.

[164] See Palmer (1991).

[165] See Braeutigam (1993).

produces a regulated service and may, with the regulator's permission, also produce an unregulated service. The firm is privately informed about its production costs. The regulator values the welfare of consumers of the regulated service more than he values the welfare of consumers of the unregulated service. In this setting, the regulator will optimally restrict the firm's participation in the unregulated market severely when the firm claims to have high costs, but will implement less severe output distortions in the regulated market. This policy serves to mitigate the firm's incentive to exaggerate its production costs without implementing substantial output distortions in the regulated market where the regulator is particularly averse to such distortions because of their impact on the welfare of consumers of the regulated service.[166]

3.6. Conclusions

The simple, practical regulatory policies reviewed in this section complement the optimal regulatory policies reviewed in Section 2. The practical policies provide insight about the gains that regulation can secure even when the regulator's knowledge of the regulated industry is extremely limited. The optimal policies provide further insight about how a regulator can employ any additional information that he may gain about the regulatory environment to refine and improve upon simple regulatory plans.

The analyses of optimal and practical regulatory policies together provide at least four important observations. First, carefully designed regulatory policies often can induce the regulated firm to employ its superior information in the best interests of consumers. Although the objectives of the regulated firm typically differ from those of society at large, the two sets of objectives seldom are entirely incongruent. Consequently, Pareto gains often can be secured. Second, the Pareto gains are secured by delegating some discretion to the regulated firm. The (limited) discretion afforded the firm is the means by which it can employ its superior knowledge to secure Pareto gains. The extent of the discretion that is optimally afforded the firm will depend upon both the congruity of the preferences of the regulator and the firm and the nature and extent of the prevailing information asymmetry.

Third, it generally is not costless to induce the firm to employ its superior information in the best interests of consumers. The firm typically will command rent from its superior knowledge of the regulatory environment. Although the regulator may place little or no value on the firm's rent, any attempt to preclude all rent can eliminate large potential gains for consumers. Consequently, the regulator may further the interests of consumers by credibly promising not to usurp all of the firm's rent. Fourth, the regulator's ability to achieve his objectives is influenced significantly by the instruments at his

166 See Anton and Gertler (1988). Lewis and Sappington (1989c) also demonstrate how a regulator can secure gains for regulated customers by limiting the firm's participation in an unregulated market severely when it claims to have high operating costs in the regulated market. Sappington (2003) examines the optimal design of diversification rules to prevent a regulated firm from devoting an excessive portion of its limited resources to reducing its operating costs in diversified markets.

disposal. The regulator with fewer instruments than objectives typically will be unable to achieve all of his objectives, regardless of how well informed he is about the regulatory environment. Of course, limited information compounds the problems associated with limited instruments.

This fourth observation, regarding the instruments available to the regulator, is also relevant to the discussion in Section 4. The discussion there explains how a regulator can employ another instrument – potential or actual competition – to discipline the regulated firm and increase social welfare.

4. Optimal regulation with multiple firms

Even though regulation often is implemented in monopoly settings, it frequently is implemented in other settings as well. Consequently, the design of regulatory policy often must account for the influence of competitive forces. The primary purpose of this section is to consider how competitive forces can be harnessed to improve regulatory policy. This section also considers how the presence of competition can complicate the design of regulatory policy.

Competition has many potential benefits.[167] The present discussion focuses on two of these benefits: the *rent-reducing benefit* and the *sampling benefit*. In a competitive setting, the regulator may be able to terminate operations with a supplier who claims to have high costs because the regulator can secure output from an alternative supplier. Consequently, firms may have limited leeway to misrepresent their private information, and so may command less rent from their private information. This is the rent-reducing benefit of competition. The sampling benefit of competition emerges because, as the number of potential suppliers increases, the chosen supplier is more likely to be a particularly capable one. Together, the sampling and rent-reducing benefits of competition can help the regulator to identify a capable supplier and to limit the rent that accrues to the supplier.

The analysis of these benefits of competition and other benefits of competition begins in Section 4.1, which examines the design of yardstick competition. Under yardstick competition, a monopoly supplier in one jurisdiction is disciplined by comparing its activities to the activities of monopolists that operate in other jurisdictions. Section 4.2 analyzes the optimal design of competition *for* a market when competition *in* the market is precluded by scale economies and when yardstick competition is precluded by the absence of comparable operations in other jurisdictions. Section 4.3 examines how the presence of unregulated rivals affects the design of regulatory policy for a dominant supplier.

In contrast to Sections 4.1 through 4.3, which take the industry structure as given and beyond the regulator's control, Sections 4.4 and 4.5 examine the optimal structuring of

[167] See Vickers (1995b), for instance, for a survey.

a regulated industry. Section 4.4 analyzes the number of firms that a regulator should authorize to produce a single product. Section 4.5 extends this analysis to settings where there are multiple regulated products, and the regulator can determine which firm supplies which product. Integration of production activities (i.e., choosing a single firm to supply all products) can provide a rent-reducing benefit, unless there is strong correlation between the costs of supplying the various services or unless the products are close substitutes. Section 4.6 considers the additional complications that arise when the quality of the regulated products delivered by multiple (actual or potential) suppliers is difficult for the regulator and/or for consumers to discern. Section 4.7 summarizes key conclusions regarding the interplay between regulation and competition.

4.1. Yardstick competition

In some settings, scale economies render operation by two or more firms within the same market prohibitively costly. However, even when direct competition among firms is not feasible within a market, a regulator may still be able to harness competitive forces to discipline a monopoly provider. He may do so by basing each firm's compensation on its performance (or report) relative to the performance (or reports) of firms that operate in other markets. When the firms are known to operate in similar environments, yardstick competition can produce a powerful rent-reducing benefit. The benefit can be so pronounced as to ensure the full-information outcome. We develop this conclusion in two distinct settings, which we refer to as the "yardstick performance" setting and the "yardstick reporting" setting. The sampling benefit of competition does not arise in either of these settings because, by assumption, there is only a single firm that is available to operate in each market.

4.1.1. Yardstick performance setting

To illustrate the potential value of yardstick competition, consider the following simple yardstick performance setting.[168] Suppose there are n identical and independent markets, each served by a separate monopolist. The local monopolists all face the same demand curve, $Q(p)$, and have identical opportunities to reduce marginal costs. Specifically, suppose $F(c)$ is the fixed cost that a firm must incur to achieve marginal cost c. The regulator is assumed to have no knowledge of the functional form of either $Q(\cdot)$ or $F(\cdot)$. However, the regulator can observe a firm's realized marginal cost of production c_i and its cost-reducing expenditures F_i in each market $i = 1, \ldots, n$. The regulator specifies the price p_i that firm i must set and the transfer payment T_i that will be awarded to firm i. The regulator seeks to maximize the total surplus generated in the n markets, while ensuring that each producer makes non-negative profit. After observing the prices and transfer payments specified by the regulator, the firms choose cost-reducing expenditure levels simultaneously and independently. Each firm acts to maximize its profit,

[168] The following discussion is based on Shleifer (1985).

taking as given the predicted actions of the other firms. Collusion is assumed not to occur in this yardstick performance setting.

Proposition 12 reveals how, despite his limited knowledge, the regulator can exploit the symmetry of the environments to achieve the full-information outcome. In the full-information outcome, the price in each market equals the realized marginal cost of production ($p_i = c_i$) and each firm undertakes cost-reducing expenditures up to the point at which the marginal expenditure and the associated marginal reduction in operating costs are equal (i.e., $Q(c_i) + F'(c_i) = 0$, as in expression (11) above).

PROPOSITION 12. *The regulator can ensure the full-information outcome as the unique symmetric Nash equilibrium among the monopolists in the yardstick performance setting by setting each firm's price equal to the average of the marginal costs of the other firms and providing a transfer payment to each firm equal to the average cost-reducing expenditure of the other firms. Formally,*

$$p_i = \frac{1}{n-1}\sum_{j \neq i} c_j; \qquad T_i = \frac{1}{n-1}\sum_{j \neq i} F_j \quad \textit{for } i = 1, \ldots, n.$$

Since each firm's compensation is independent of its own actions under the reward structure described in Proposition 12, each firm acts to minimize its own production costs ($c_i Q(p_i) + F(c_i)$). The requirement to price at the average realized marginal cost of other producers then ensures prices that maximize total surplus. The authorized prices and transfer payments provide zero rent to all producers in this symmetric setting.

Proposition 12 illustrates vividly the potential gains from yardstick competition. Even when the regulator has little knowledge of the operating environment in each of the symmetric markets, he is able to ensure the ideal outcome in all markets.[169] In principle, corresponding results could be achieved if the producers faced different operating environments. In this case, though, the regulator would require detailed knowledge of the differences in the environments in order to ensure the full-information outcome.[170] Failure to adjust adequately for innate differences in operating environments could lead to financial hardship for some firms, significant rents for others, and suboptimal levels of cost-reducing expenditures.[171]

[169] Notice, in particular, that the regulator does not have well defined Bayesian prior beliefs about the functional form of each firm's technological capabilities, just as in the non-Bayesian models of regulation reviewed in Section 3. The regulator's ability to ensure the full-information outcome here is reminiscent of his ability to induce Ramsey prices with the lagged expenditure policy discussed in Section 3.2.1. There, the repeated observation of the performance of a single myopic monopolist in a stationary environment plays the same role that the observation of the performance of multiple monopolists in symmetric environments plays in the current context.

[170] See Shleifer (1985) for a discussion of how the regulatory policy might be modified when different firms produce in different operating environments.

[171] See, for example, Nalebuff and Stiglitz (1983).

A crucial simplifying feature of the yardstick performance setting is that the firms face no uncertainty.[172] If uncertainty is introduced into the production functions, then the full-information outcome typically is not possible when firms are risk averse. This is because the regulator must consider the firms' aversion to risk when determining the optimal power of the incentive scheme (as discussed in Section 2.6.2). The policy proposed in Proposition 12 is high powered and would expose risk-averse firms to excessive risk. Nevertheless, even when there is uncertainty and when firms are risk averse, it is generally optimal to condition each firm's reward on the performance of other firms, thereby incorporating yardstick competition to some degree.[173]

Despite the pronounced gains it can secure in some settings, yardstick competition can discourage innovative activity when spillovers are present or when the regulator's commitment powers are limited. To illustrate, suppose the cost-reducing expenditure of each firm in the yardstick performance setting serves to reduce both its own costs and (perhaps to a lesser extent) the costs of other firms. Then a reward structure like the one described in Proposition 12 will not induce the full-information outcome because it does not reward each firm fully for the beneficial impacts of its expenditures on the costs of other firms. Indeed, the price a firm is permitted to charge would decline as its cost-reducing expenditure increased, since the increased expenditure would reduce the operating costs of the other firms. More generally, when externalities of this sort are present and when the regulator cannot commit in advance to limit his use of yardstick regulation to extract rent from the regulated firms, social welfare can be lower when the regulator is empowered to employ yardstick regulation than when he is precluded from doing so.[174]

4.1.2. Yardstick reporting setting

Yardstick competition also can admit a powerful rent-reducing benefit simply by comparing the cost reports of actual or potential competitors. To illustrate this fact, consider the following yardstick reporting setting, which parallels the setting considered in Section 2.4.1.[175] There are two firms, A and B, that operate in correlated environments. Firm A has exogenous marginal cost $c^A \in \{c_L^A, c_H^A\}$ and fixed cost F^A. Firm B has

[172] The yardstick performance setting also abstracts from potential collusion among producers. Potters et al. (2004) present an experimental study of the extent of collusion under different yardstick competition policies.
[173] See Mookherjee (1984) for an analysis of the moral hazard problem with several agents. Mookherjee shows that, except in the special case where the uncertainty faced by the agents is perfectly correlated, the full-information outcome is not possible when agents are risk averse. He also shows that the optimal incentive scheme for one agent should depend on the performance of other agents whenever uncertainty is correlated. Also see Armstrong, Cowan and Vickers (1994, Section 3.4) for a simplified analysis in which regulatory policy is restricted to linear schemes.
[174] Dalen (1998) and Sobel (1999) prove this observation. Meyer and Vickers (1997) provide related insights in their analysis of implicit rather than explicit relative performance comparisons.
[175] This discussion is based on the analysis in Demski and Sappington (1984) and Crémer and McLean (1985).

marginal cost $c^B \in \{c^B_L, c^B_H\}$ and fixed cost F^B. Fixed costs are common knowledge, but each firm is privately informed about its realized marginal cost.

Initially, suppose that firm B can be relied upon to report its cost truthfully, and consider the optimal policy towards firm A. Let ϕ^A_i denote the probability that firm B has a low-cost realization c^B_L when firm A's marginal cost is c^A_i, for $i = L, H$. To capture the fact that the two firms operate in correlated environments, assume $\phi^A_L > \phi^A_H$. Just as in Section 2.4.1, the regulator can ensure marginal-cost pricing for firm A without ceding any rent if there are no bounds on the penalties that can be imposed on the risk-neutral firm. He can do so by conditioning the transfer payment to firm A on its report of its own cost and on the cost report of firm B.

Specifically, let T^A_{ij} be the transfer payment to firm A when it claims its cost is c^A_i and when firm B claims its cost to be c^B_j. If firm A claims to have a high cost, it is permitted to charge the unit price $p^A_H = c^A_H$. In addition, firm A receives a generous transfer payment when firm B also claims to have high costs, but is penalized when firm B claims to have low costs. These transfer payments can be structured to provide an expected transfer of F^A to firm A when its marginal cost is indeed c^A_H, so that

$$\phi^A_H T^A_{HL} + \left(1 - \phi^A_H\right) T^A_{HH} = F^A.$$

At the same time, the payments can be structured to provide an expected return to firm A when it has low costs that is sufficiently far below F^A that it eliminates any rent firm A might anticipate from being able to set the relatively high price ($p^A_H = c^A_H$), so that

$$\phi^A_L T^A_{HL} + \left(1 - \phi^A_L\right) T^A_{HH} \ll F^A.$$

The transfers T^A_{HL} and T^A_{HH} can always be set to satisfy this pair of expressions except when the costs of the two firms are independently distributed ($\phi^A_L = \phi^A_H$). When firm A reports it has low cost, it is simply offered its (deterministic) full information contract, with price equal to c^A_L and transfer payment equal to F^A. Consequently, provided that firm B reports its cost truthfully, the full-information outcome can be implemented for firm A with this pair of contracts. Firm B's cost report serves precisely the same role that the audit did in Section 2.4.1.

Of course, an identical argument can be applied to the regulation of firm B. In particular, if firm A can be induced to report its cost truthfully, then the full-information outcome can be implemented for firm B. Consequently, a yardstick reporting policy can implement the full-information outcome in both markets as a Nash equilibrium. Thus, even a very limited correlation among firms' costs can constitute a powerful regulatory instrument when feasible payments to firms are not restricted and when firms are risk neutral. This is because a firm with a low cost knows that other firms are also likely to have low costs. Consequently, cost exaggeration poses considerable risk of a severe penalty.

When the firms' costs are not highly correlated, extreme penalties may be required to eliminate a firm's unilateral incentive for cost exaggeration. Just as in Section 2.4.1, this can be problematic if firms are risk averse or if feasible payoffs to firms are bounded.[176] Another potential complication with a yardstick reporting policy of this type is that it might encourage the firms to coordinate their behavior. Although there is an equilibrium where the two firms truthfully report their private cost information, other equilibria can arise in which the firms systematically exaggerate their costs, leading to high prices and rent for the firms. More generally, when firms are rewarded according to how their performance or their reports compare to the performance or reports of their peers, the firms typically can coordinate their actions or reports and thereby limit the regulator's ability to implement effective yardstick competition.[177]

4.2. Awarding a monopoly franchise

Yardstick regulation relies upon the operation of monopolists in distinct markets. In contrast, franchise bidding creates competition among multiple potential suppliers for the right to serve as a monopolist in a single market.[178] Such competition can promote both sampling and rent-reducing benefits.

4.2.1. A static model

To illustrate how a regulator might employ franchise bidding to discipline a monopoly supplier, consider the following setting based on the Baron–Myerson model described in Section 2.3.1. Suppose there are $N \geqslant 1$ firms that are qualified to serve as a monopoly provider in a particular market.[179] Each firm has either low marginal cost (c_L) or high marginal cost (c_H). Let ϕ denote the probability that a given firm has a low-cost realization, and suppose the costs of the N firms are distributed independently.[180] The firm

[176] Demski and Sappington (1984) analyze a setting where firms are risk averse. Demski, Sappington and Spiller (1988), Dana (1993) and Lockwood (1995), among others, consider settings where feasible rewards and penalties are bounded.

[177] Ma, Moore and Turnbull (1988), Glover (1994), and Kerschbamer (1994) show how reward structures can be modified in adverse selection settings to rule out undesired equilibria in which firms systematically misreport their private cost information. Laffont and Martimort (1997) and Tangerås (2002) analyze the additional insights that arise when regulated firms are able to coordinate their actions explicitly. For instance, Tangerås (2002) shows that the value of yardstick competition becomes negligible as the firms' private cost information becomes perfectly correlated.

[178] Demsetz (1968) provides a pioneering discussion of the merits of franchise bidding.

[179] See Kjerstad and Vagstad (2000) for an analysis of the case where the number of participating bidders depends on the expected rents from the auction. Taylor (1995), Fullerton and McAfee (1999), and Che and Gale (2003) demonstrate the merits of limiting the number of firms that are permitted to compete for the right to supply a product. The entry restrictions increase the likelihood that any particular firm will be selected to produce, and thereby increase each firm's incentive to incur sunk development costs that improve its performance.

[180] The regulator can achieve the full-information outcome in this setting if the firms' costs are correlated by making use of the yardstick reporting mechanism discussed in Section 4.1.2.

that actually produces incurs the known fixed cost F. When F is sufficiently large, the regulator will optimally authorize the operation of only one producer.[181]

The optimal regulatory policy in this setting is readily shown to take the following form. After the regulator announces the terms of the regulatory policy, the firms simultaneously announce their cost realizations. If at least one firm claims to have low costs, one of these firms is selected at random to serve as the monopoly supplier. If all N firms report high costs, one of the firms is selected at random to be the monopoly supplier. The regulatory policy specifies that when a firm is selected to produce after reporting cost c_i, the firm must charge price p_i for its product and receive a transfer payment T_i from the regulator.[182] When a firm that truthfully announces cost c_i is selected to produce, it will receive rent $R_i = Q(p_i)(p_i - c_i) - F + T_i$. A firm that announces cost c_i will be selected to produce with probability ρ_i. In the equilibrium where all firms announce their costs truthfully (which can be considered without loss of generality if there is no collusion between firms), a firm that announces it has high costs will only win the contract when all other firms have high costs, and in that case only with probability $1/N$. Therefore,

$$\rho_H = \frac{(1-\phi)^{N-1}}{N}$$

is the probability that a firm that announces it has high costs will win the auction. Similarly, if a firm announces it has low costs, it will win the contest with the (if $N > 1$) higher probability[183]

$$\rho_L = \frac{1-(1-\phi)^N}{N\phi}.$$

Therefore, taking into account its probability of winning, the equilibrium expected rent of a firm with cost c_i is $\rho_i R_i$.

Now consider the incentive compatibility constraints that must be satisfied. As with expression (4), if a low-cost firm claims to have high costs and wins the contest, it will earn rent $R_H + \Delta^c Q(p_H)$. However, cost exaggeration reduces the equilibrium probability of winning the franchise from ρ_L to ρ_H. Consequently, a truthful report of low cost is ensured if $\rho_L R_L \geqslant \rho_H[R_H + \Delta^c Q(p_H)]$, or

$$R_L \geqslant \frac{\rho_H}{\rho_L}\big[R_H + \Delta^c Q(p_H)\big]. \tag{88}$$

Comparing expression (88) with expression (4), the corresponding constraint when there is only one potential supplier, it is apparent that competition relaxes the rele-

[181] The possibility of simultaneous production by multiple producers is considered below in Section 4.4, as is the possibility of an endogenous number of active producers.

[182] In principle, p_i and T_i might vary with the costs reported by the firms that are not selected to operate. However, such variation provides no strict gains when the costs of all potential suppliers are independent.

[183] For instance, see Lemma 1 in Armstrong (2000).

vant incentive compatibility constraint.[184] This is the rent-reducing benefit of franchise bidding.

As in expression (15), social welfare when a firm with cost c_i is selected to produce is $w_i(p_i) - (1-\alpha)R_i$, where $w_i(p_i)$ is total surplus when price is p_i and $\alpha \leqslant 1$ is the weight the regulator places on rent. Since the probability that a low-cost firm is selected to produce is $1-(1-\phi)^N$, total expected welfare is

$$W = \left(1-(1-\phi)^N\right)\left\{w_L(p_L) - (1-\alpha)R_L\right\} \\ + (1-\phi)^N\left\{w_H(p_H) - (1-\alpha)R_H\right\}. \tag{89}$$

Comparing expression (89) with expression (16), the corresponding expression when there is only one potential producer, reveals the sampling benefit of competition: the probability that the monopoly producer has low cost increases.

Standard arguments show that $R_H = 0$ and $p_L = c_L$ under the optimal policy. Also, the incentive constraint (88) will bind, and so p_H is chosen to maximize

$$(1-\phi)^N w_H(\cdot) - \left(1-(1-\phi)^N\right)\frac{\rho_H}{\rho_L}(1-\alpha)\Delta^c Q(\cdot).$$

Therefore, the price charged by the high-cost firm is

$$p_H = c_H + \frac{\phi}{1-\phi}(1-\alpha)\Delta^c,$$

which does not depend on N, and is exactly the optimal price specified in expression (5) that prevails in the absence of competition for the market.

It may be surprising that (conditional on the realized cost) the prices ultimately charged by the selected supplier do not vary with the number of firms that compete to serve as the monopoly supplier.[185] This invariance holds because two conflicting effects offset each other. The first effect arises because a low-cost firm that faces many competitors for the franchise is less tempted to exaggerate its cost, since the exaggeration reduces the probability (from ρ_L to ρ_H) that it will be selected to operate the franchise. Consequently, a smaller output distortion for a high-cost firm is needed to deter cost exaggeration, and so p_H can be reduced toward c_H. The second effect arises because the likelihood that a low-cost firm will be awarded the franchise increases as N increases. Therefore, it becomes more important to reduce the rent of the low-cost firm by raising p_H above c_H. These two effects turn out to offset each other exactly in this setting with risk-neutral firms and independently distributed costs.

Expression (88) reveals that the equilibrium rent of a low-cost firm that wins the contest is $R_L = \frac{\rho_H}{\rho_L}\Delta^c Q(p_H)$. Since ρ_H/ρ_L is decreasing in the number of bidders and

[184] As usual, the only binding incentive compatibility constraint is the one that ensures the low-cost firm will not exaggerate its cost.

[185] This result is not an artifact of the particular framework we use here (involving exogenous costs and binary realizations). Laffont and Tirole (1987) term the result the 'separation property'.

the high-cost price p_H is independent of the number of bidders, this rent decreases with the number of bidders.[186] Furthermore, since the probability that a low-cost firm wins is $[1 - (1 - \phi)^N]$, the aggregate expected rent of all bidders is

$$\left[1 - (1 - \phi)^N\right] \frac{\rho_H}{\rho_L} \Delta^c Q(p_H) = \phi(1 - \phi)^{N-1} \Delta^c Q(p_H). \tag{90}$$

This expected industry rent is decreasing in N. These key features of the optimal regulatory policy in this setting are summarized in Proposition 13.[187]

PROPOSITION 13. *The optimal franchise auction in this static setting with independent costs has the following features*:

(i) *The franchise is awarded to the firm with the lowest cost.*
(ii) *A high-cost firm makes zero rent.*
(iii) *The rent enjoyed by a low-cost firm that wins the contest decreases as the number of bidders increases.*
(iv) *The total expected rent of the industry decreases as the number of bidders increases.*
(v) *The prices that the winning firm charges do not depend on the number of bidders, and are the optimal prices in the single-firm setting, as specified in expression (5).*

This static analysis of franchise auctions has assumed that all potential operators are identical ex ante. When some operators are known to have higher expected costs than others, it can be advantageous to favor these operators by awarding the franchise to them with higher probability than it is awarded to operators with lower expected cost, ceteris paribus. Doing so can induce the operators with lower expected costs to bid more aggressively than they would in the absence of such handicapping.[188,189] Because such a policy may not award the franchise to the least-cost supplier, the policy intentionally sacrifices some productive efficiency in order to reduce the rent enjoyed by low-cost firms.

[186] When potential operators have limited resources, more capable operators cannot necessarily outbid their less capable rivals. Consequently, Lewis and Sappington (2000) show that the potential operators may resort instead to sharing larger fractions of realized profit with consumers. See Che and Gale (1998, 2000) for related analyses.

[187] Parallel results are obtained by Riordan and Sappington (1987a), Laffont and Tirole (1987), and McAfee and McMillan (1987b). Riordan and Sappington (1987a) analyze a model where the firm has only imperfect information about its eventual cost at the time of bidding. The other two studies examine settings where realized production costs are endogenous and observable.

[188] For instance, see the discussion in McAfee and McMillan (1987a, Section VII).

[189] We have not discussed the possibility of collusion between the regulator and one or more bidders, which is another kind of "favoritism". For discussions of this point, see Laffont and Tirole (1991a) and Celentani and Ganuza (2002).

4.2.2. Dynamic considerations

Although franchise bidding admits the rent-reducing and sampling benefits of competition, it is not without its potential drawbacks. These drawbacks include the following three.[190] First, it may be difficult to specify fully all relevant dimensions of performance, particularly if the franchise period is long. Therefore, actual performance may fall short of ideal performance on many dimensions, as the firm employs unavoidable contractual incompleteness to its own strategic advantage. Second, a franchise operator may be reluctant to incur sunk investment costs if there is a substantial chance that its tenure will end before the full value of the investment can be recovered. Consequently, the supplier may not operate with the least-cost technology. Third, incumbency advantages (such as superior knowledge of demand and cost conditions or substantial consumer loyalty) can limit the intensity of future competition for the right to serve as the franchise operator, as new potential operators perceive their chances of winning the contract on profitable terms to be minimal.[191]

To overcome the first of these potential drawbacks (contractual incompleteness), it may be optimal to award the monopoly franchise for a relatively short period of time. In contrast, the second potential drawback (limited investment incentives) may be best mitigated by implementing a relatively long franchise period, thereby providing a relatively long period of time over which the incumbent can benefit from its investments. To alleviate the tension introduced by these two counteracting effects, it may be optimal to award a franchise contract for a relatively short period of time, but to bias subsequent auctions in favor of the incumbent. Of course, such a policy can aggravate the third potential drawback to franchise bidding (incumbency advantages).

Although biasing franchise renewal auctions in favor of the incumbent can aggravate the potential problems caused by incumbency advantages, such biasing can be optimal when non-contractible investments by the incumbent reduce operating costs or enhance product quality substantially and when the benefits of these investments flow naturally to future franchise operators. Increasing the likelihood that the incumbent will be selected to operate the franchise in the future can increase the incumbent's expected return from such transferable, sunk investments. Consequently, such a bias can enhance incentives for the incumbent to undertake these valuable investments.[192] By contrast, when its investments are not transferable to rivals, the incumbent has stronger incentives to undertake such investments. In such a case, because the incumbent is expected to have

[190] Williamson (1976) discusses these potential drawbacks in more detail. Prager (1989), Zupan (1989b, 1989a) and Otsuka (1997) assess the extent to which these potential problems arise in practice.

[191] If incumbent suppliers acquire privileged information about the profitability of serving the franchise area, non-incumbent potential suppliers may not bid aggressively for the right to serve the franchise area, for fear of winning the franchise precisely when they have over-estimated its value.

[192] An examination of the optimal policy to motivate transferable investment by an incumbent would naturally include a study of the optimal length of the monopoly franchise, as discussed in Section 3.2.3.

lower operating costs than its rivals in subsequent auctions, it can be optimal to bias the subsequent auctions against the incumbent.[193]

Second sourcing in procurement settings is similar to franchise renewal in regulatory settings. Under second sourcing, the regulator may transfer operating rights from an incumbent to an alternative supplier. The second source might be a firm that presently serves other markets, or it might be a potential supplier that does not presently operate elsewhere. Second sourcing can increase welfare: (1) by shifting production from the incumbent to the second source when the latter has lower operating costs than the former (the sampling benefit); and (2) by reducing the rent the supplier secures from its privileged knowledge of its operating environment. This rent-reducing effect can arise for two reasons. First, as reflected in expression (88) above, the incumbent will be less inclined to exaggerate its operating costs when the probability that it is permitted to operate declines as its reported costs increase.[194] Second, when the incumbent's production technology can be transferred to the second source, the technology may generate less rent for the second source than it does for the incumbent. This will be the case if cost variation under the incumbent's technology is less sensitive to variations in the innate capability of the second source than it is to the corresponding variation in the incumbent's ability.[195]

When the operating costs of the incumbent and the second source are correlated, the optimal second-sourcing policy can share some features of the auditing and yardstick policies described in Sections 2.4.1 and 4.1.2. In particular, an incumbent that reports high cost can be punished (by terminating its production rights) when the second source reports low cost. In contrast, the incumbent can be rewarded when the second source corroborates the incumbent's report by reporting high cost also. However, an optimal second sourcing policy differs from an optimal auditing policy in at least two respects. First, cost reports by the second source are endogenous and are affected by the prevailing regulatory policy. Second, a second source enables the regulator to alter the identity of the producer while an audit in a monopoly setting does not change the producer's identity. These differences can lead the regulator to solicit a costly report from the second source more or less frequently than he will undertake an equally costly audit, and to set different prices in the regulated industry in response to identical reports from an audit and a second source. To best limit the rent of the incumbent, it can be optimal

[193] Laffont and Tirole (1988b) analyze these effects in detail. See also Luton and McAfee (1986) for a model without investment.

[194] Sen (1996) demonstrates the useful role that the threat of termination can play in adverse selection settings. He shows that when a regulator can credibly threaten to replace an incumbent with a second source, the quantity distortions that are implemented to limit information rents may be reduced. Also see Dasgupta and Spulber (1989/1990). Anton and Yao (1987) demonstrate the benefits of being able to shift production to a second source even when doing so can increase industry costs by foregoing valuable learning economies.

[195] For example, when it operates with the incumbent's technology, the second source's marginal cost of production may be a weighted average of its own innate cost and that of the incumbent. See Stole (1994).

to use the second source even when it is known to have a higher cost than the incumbent.[196]

Although second sourcing may increase welfare, second sourcing (like auditing) does not necessarily do so when the regulator has limited commitment powers. Second sourcing can reduce welfare by enabling the regulator to limit severely the rent the incumbent earns when its operating costs are low. When it anticipates little or no rent from realizing low production costs, the incumbent will not deliver substantial unobservable cost-reducing effort. Therefore, in settings where substantial cost-reducing effort is desirable and where limited commitment powers force the regulator to implement the policy that is best for consumers after the incumbent has delivered its cost-reducing effort, welfare can be higher when second sourcing is not possible. In essence, eliminating the possibility of second sourcing helps to restore some of the commitment power that is needed to motivate cost-reducing effort.[197]

4.3. Regulation with unregulated competitive suppliers

Situations often arise where a dominant firm and a number of smaller firms serve the market simultaneously, and the regulator only controls directly the activities of the dominant firm.[198] In these settings, the presence of alternative unregulated suppliers can affect both the optimal regulation of the dominant firm and overall welfare in a variety of ways. Competition can enhance or reduce welfare. While competition can introduce the rent-reducing and sampling benefits, unregulated competitors may undermine socially desirable tariffs that have been imposed on the regulated firm.

To analyze these effects formally, consider the following simple example which extends the Baron–Myerson model summarized in Section 2.3.1. Suppose the dominant firm's marginal cost is either low c_L or high c_H. In the absence of competition, the optimal regulatory policy would be as specified in Proposition 1. Suppose now there are a large number of rivals, each of which supplies exactly the same product as the dominant firm and each of which has the (known) unit cost of supply, c^R. Competition within this "competitive fringe" ensures the fringe always offers the product at price c^R. (For simplicity, we abstract from fixed costs of production for the fringe and the regulated dominant firm.)

There are four cases of interest, depending on the level of the fringe's cost c^R. First, suppose $c^R < c_L$. The fringe will increase welfare in this case because the industry

[196] See Demski, Sappington and Spiller (1987) for details.

[197] See Riordan and Sappington (1989) for a formal analysis of this effect. Notice that the decision to eliminate a second source here serves much the same role that favoring the incumbent supplier plays in the franchise bidding setting analyzed by Laffont and Tirole (1988b). Of course, as Rob (1986) and Stole (1994) demonstrate, if the regulator's commitment powers are unimpeded, second sourcing typically will improve welfare even when substantial unobservable cost-reducing effort is socially desirable.

[198] In contrast, in the models of second sourcing discussed in the previous section, the regulator could choose when to allow entry, and on what terms.

price and production costs are always lower when the fringe is active. Second, suppose $c_L < c^R < c_H$. Here too, the fringe increases welfare. The optimal regulatory policy in this case requires the dominant firm to set the price $p = c_L$. The firm will reject this contract if its cost is high, in which case the market is served by the fringe. This policy ensures the full-information outcome: the least-cost provider supplies the market, price is equal to marginal cost, and no firm receives any rent. Thus, the competitive fringe provides both a sampling and a rent-reducing benefit in this setting. The sampling benefit arises because the fringe supplies the market at lower cost than can the high-cost dominant firm. The rent-reducing effect arises because the low-cost dominant firm has no freedom to exaggerate its costs. Third, suppose $c^R > p_H$, where p_H is given in expression (5). In this case, the fringe has no impact on regulatory policy. The fringe's cost is so high that it cannot undercut even the inflated price of the high-cost firm, and so the policy recorded in Proposition 1 is again optimal.

The final (and most interesting) case arises when $c_H < c^R < p_H$. In this case, the marginal cost of the fringe always exceeds the marginal cost of the dominant firm. However, the cost disadvantage of the fringe is sufficiently small that it can profitably undercut the price (p_H) that the high-cost dominant firm is optimally induced to set in the absence of competition. Therefore, the presence of the fringe admits two regulatory responses: (i) reduce the regulated price from p_H to c^R for the high-cost dominant firm, thereby precluding profitable operation by the fringe; or (ii) allow the fringe to supply the entire market (at price c^R) when the dominant firm has high cost. Policy (ii) is implemented by requiring the dominant firm to charge the price equal to c_L if it wishes to supply the market.

Policy (i) offers the potential advantages of ensuring production by the least-cost supplier and moving price closer to marginal cost when the dominant firm has high costs. However, these potential gains are more than offset by the additional rent that policy (i) affords the dominant firm. Recall from Equations (28) and (29) that once the rent of the dominant firm is accounted for, expected welfare is the welfare derived from the setting in which the regulator is fully informed about the firm's cost but where the high cost is inflated to p_H. Because the fringe has a lower marginal cost than the adjusted cost of the high-cost dominant firm ($c^R < p_H$), expected welfare is higher when the fringe operates in place of the high-cost dominant firm.[199]

Proposition 14 summarizes these observations.

PROPOSITION 14. *Consumer surplus and welfare are higher, and the rent of the dominant firm is lower, in the competitive fringe setting than in the corresponding setting where the fringe does not operate.*

Notice that competition does not undermine socially desirable prices or otherwise reduce welfare in this simple setting. The same is true in similar settings, where the

[199] This same logic explains why a regulator might favor a less efficient bidder in a franchise auction, as discussed in Section 4.2.

fringe's cost is uncertain and may be correlated with the dominant firm's cost.[200,201] However, competition can reduce welfare in some settings. It might do so, for example, by admitting "cream-skimming", which occurs when competitors attempt to attract only the most profitable customers, leaving the incumbent regulated supplier to serve the less profitable (and potentially unprofitable) customers. To illustrate this possibility, consider the following simple setting. Suppose the incumbent regulated firm has no private information about its cost of operation. The friction in this setting arises because (in contrast to most of the other settings considered in this survey) there is a social cost of public funds $\Lambda > 0$.[202] (See Section 2.1 above.) The firm offers n products at prices $\mathbf{p} = (p_1, \ldots, p_n)$. At these prices, the firm's profit is $\pi(\mathbf{p})$ and consumer surplus is $v(\mathbf{p})$. In this setting, as in expression (1), welfare is $v(\mathbf{p}) + (1 + \Lambda)\pi(\mathbf{p})$. In the absence of competition, optimal (Ramsey) prices $\mathbf{p}^*$ maximize this expression.

Now suppose there is a competitive fringe that supplies a single product (product i) at price (and cost) equal to c_i^R. If $c_i^R > p_i^*$, the fringe does not interfere with the Ramsey prices. However, if $c_i^R < p_i^*$, the fringe will undercut the Ramsey price for product i. The lower price could increase welfare if the fringe's cost is sufficiently small relative to the corresponding cost of the regulated firm. However, if the fringe's cost advantage is sufficiently limited, welfare will decline. This is most evident when the two marginal costs are identical. In this case, the fringe does not reduce industry operating costs, but its presence forces a price for product i below the Ramsey price, p_i^*. When the fringe has higher costs than the regulated firm but can still operate profitably at price p_i^*, the operation of the fringe will both raise industry costs and divert prices from their Ramsey levels. Consequently, an unregulated competitive fringe can limit the options available to the regulator without offering offsetting benefits, such as those that arise from the rent-reducing or sampling benefits of competition.[203,204]

200 See Caillaud (1990). Caillaud shows that when the costs of the regulated firm and the fringe are positively correlated, smaller output distortions will be implemented when the competitive fringe is present. When costs are positively correlated, the regulated firm is less tempted to exaggerate costs, ceteris paribus, because it anticipates that the fringe will have low cost when the regulated firm does. Consequently, the reduced output that the regulated firm will be authorized to produce when it exaggerates cost will induce the fringe to supply a particularly large output level, resulting in a low market price and low profit for the regulated firm. The regulator responds to the firm's reduced incentive for cost exaggeration by imposing smaller output distortions.

201 Biglaiser and Ma (1995) show that when firms supply differentiated products and have superior knowledge of market demand, the presence of an unregulated producer can have different qualitative effects on optimal regulatory policy. Prices can be distorted above or below marginal cost, in part to induce a preferred allocation of customers among producers.

202 If $\Lambda = 0$ the regulator could ensure the ideal full-information outcome simply by requiring marginal-cost prices and delivering the transfer required to ensure the firm's participation. Competition would be beneficial in such a setting.

203 Baumol, Bailey and Willig (1977) and Baumol, Panzar and Willig (1982) identify (restrictive) conditions under which Ramsey prices are not vulnerable to such competitive entry.

204 Laffont and Tirole (1990b) analyze a variant of this model that involves second-degree price discrimination. There are two groups of consumers, high- and low-volume users, and the fringe has a technology that is

Such undesirable entry also can occur when the regulator has distributional objectives, and favors the welfare of one group of consumers over another.[205] (For instance, telecommunications regulators often try to keep basic local service rates low, but allow relatively high rates for long distance and international calls.) The relatively high prices that the regulator would like to set on certain services may enable competitors to provide the services profitably, even if they have higher production costs than the regulated firm. Consequently, unfettered competition can both undermine Ramsey prices and prices that reflect distributional concerns, and increase industry costs.

The mark-ups of prices above marginal costs that can arise under Ramsey pricing or in the presence of distributional concerns can be viewed as taxes that consumers must pay when they purchase products from the regulated firm. These taxes are used either to fund the firm's fixed costs or to subsidize consumption by favored consumer groups. In contrast, consumers pay no such taxes when they purchase products from an unregulated competitive fringe. Consequently, the effect of unfettered competition can be to undermine the tax base. This perspective suggests an obvious solution to the problem caused by unregulated competition: consumers should be required to pay the same implicit tax whether they purchase a product from the regulated firm or from the competitors. Such a policy, which entails regulation of the competitors, can ensure that entry occurs only when the an entrant is the least-cost supplier of a product. It can also ensure that entry does not undermine policies designed to recover fixed costs most efficiently or to achieve distributional objectives. Consequently, entry will occur only when it enhances welfare.

This policy can be illustrated using the Ramsey pricing example considered just above. If the incumbent's marginal cost for product i is c_i, the implicit tax on this product under the Ramsey prices $\mathbf{p}^*$ is $t_i = p_i^* - c_i$. If consumers must also pay this tax t_i when they buy the product from a rival firm, then entry is profitable only if the entrant has lower marginal cost than the incumbent supplier.[206] Moreover, if entry does take place, the proceeds of the tax paid by the consumers of the entrant's service can

attractive only to the high-volume consumers. Competition can force the regulator to lower the tariff offered to the high-volume users in order to induce them to purchase from the regulated firm and thereby help to finance the firm's fixed costs. But when the competitive threat is severe, the reduction in the high-volume tariff may be so pronounced that low-volume customers will also find it attractive to purchase on this tariff. To deter the low-volume customers from doing so, the usage charge on the tariff is reduced below marginal cost and the fixed charge is raised just enough to leave unchanged the surplus that the tariff provides to high-volume customers. Nevertheless, the low-volume customers benefit from the opportunity to purchase on the attractive tariff that is selected by high-volume customers, and so the welfare of all users can increase in the presence of bypass competition. Aggregate welfare can decline, though, once the costs of transfer payments to the regulated firm are taken into account. See Einhorn (1987) and Curien, Jullien and Rey (1998) for further analysis of this issue.

[205] See Laffont and Tirole (1993b, ch. 6) and Riordan (2002) for discussions of this issue, and for further references.

[206] If the entrant's unit cost for this product is c_i^R, it will find it profitable to serve consumers if and only if $c_i^R + t_i \leqslant p_i^*$, i.e., if $c_i^R \leqslant c_i$.

be used to cover the incumbent's fixed costs. In practice, though, it is often impractical to levy taxes directly on the products supplied by competitors, although access charges can sometimes be employed to levy such taxes indirectly.[207]

In summary, competition can enhance welfare, in part by introducing rent-reducing and sampling benefits. However, unfettered competition also can complicate regulatory policy by undermining preferred pricing structures.

4.4. Monopoly versus oligopoly

The preceding discussion of the interaction between regulation and competition has taken as given the configuration of the regulated industry. In practice, regulators often have considerable influence over industry structure. For example, regulators typically can authorize or deny the entry of new producers into the regulated industry. This section and the next consider the optimal structuring of a regulated industry. This section analyzes the desirable number of suppliers of a single product. Section 4.5 explores multiproduct industries, and considers whether a single firm should provide all products or whether the products should be supplied by separate firms.[208]

When choosing the number of firms to operate in an industry, a fundamental trade-off often arises. Although additional suppliers can introduce important benefits (such as increased product variety and quality, and the rent-reducing and sampling benefits of competition), industry production costs can increase when production is dispersed among multiple suppliers. The trade-off between monopoly regulation and duopoly competition is illustrated in Section 4.4.1. The optimal number of industry participants is considered in Section 4.4.2 for a simple setting in which the regulator's powers are limited.

4.4.1. Regulated monopoly versus unregulated duopoly

Consider an extension of the Baron–Myerson setting of Section 2.3.1 in which the regulator can allow a rival firm to operate in the market. We will compare the performance of two regimes in this setting: the regulated monopoly regime and the unregulated duopoly regime. Monopoly regulation imposes a fixed cost $G \geqslant 0$ on society. G might include the salaries of the regulator and his staff, as well as all associated support costs for example. If the rival enters the market, the two firms engage in Bertrand price competition.

Initially, suppose the rival's marginal cost is always the same as the incumbent's marginal cost.[209] Consequently, there is no need to regulate prices if entry occurs, since competition will drive the equilibrium price to the level of the firms' marginal cost of production. Eliminating the need for regulation saves the ongoing cost of funding the regulator, G. Of course, anticipating the intense competition that will ensue, the rival

[207] See Armstrong (2001) for further discussion. Related issues are considered in Sections 5.1.2 and 5.1.3.

[208] Regulators sometimes determine whether a regulated supplier of an essential input can integrate downstream and supply a retail service in competition with other suppliers. This issue is discussed in Section 5.2.

[209] This discussion of the case of perfect cost correlation is based on Armstrong, Cowan and Vickers (1994, Section 4.1.1).

will only enter the industry if the regulator provides a subsidy that is at least as large as the rival's fixed cost of operation, F. In this setting, the regulator effectively has the opportunity to purchase an instrument (the rival's operation) that eliminates both the welfare losses that arise from asymmetric information about the incumbent firm's operating costs and the ongoing regulatory costs, G. The regulator will purchase this instrument only if the benefits it provides outweigh its cost, which is the rival's fixed operating cost F. If F is sufficiently small, the regulator will induce the rival to operate.

Of course, the costs of an incumbent supplier and a rival are unlikely to be perfectly correlated. To examine the effects of imperfect cost correlation, consider the following setting.[210] Suppose that if the regulator authorizes entry by a rival producer, the rival and incumbent produce identical products and engage in Bertrand price competition. For simplicity, suppose the two firms have the same fixed cost F of production, but their marginal costs $c \in \{c_L, c_H\}$ may differ. Each firm has the low marginal cost c_L with probability $\frac{1}{2}$. The parameter $\rho \in [\frac{1}{2}, 1]$ is the probability that the two firms have the same cost.[211] The firms' costs are uncorrelated when $\rho = \frac{1}{2}$ and perfectly correlated when $\rho = 1$. Suppose each firm knows its own marginal cost and its rival's marginal cost when it sets its price. No transfer payments to or from the firms in the industry are possible in this duopoly setting. Initially, suppose the two firms find it profitable to operate in the industry.

Bertrand price competition ensures the equilibrium price will be c_H except when both firms have low cost, which occurs with probability $\frac{1}{2}\rho$. A firm's operating profit is zero unless it has the low marginal cost and its rival has the high marginal cost. The low-cost firm in this case secures profit $\Delta^c Q(c_H)$.[212] This positive profit is realized with probability $(1-\rho)$. Consequently, expected industry profit in this unregulated duopoly setting is $(1-\rho)\Delta^c Q(c_H)$, which declines as the firms' costs become more highly correlated. Notice that the probability that the industry supplier has low marginal cost is $(1-\frac{1}{2}\rho)$. This probability decreases as ρ increases, because the sampling benefit of competition is diminished when costs are highly correlated. In contrast, the probability that the industry price will be c_L is $\frac{1}{2}\rho$, which increases as ρ increases. If α is the relative weight on industry profit in the welfare function, expected welfare in this unregulated duopoly setting is

$$\frac{1}{2}\rho v(c_L) + \left(1 - \frac{1}{2}\rho\right) v(c_H) + \alpha(1-\rho)\Delta^c Q(c_H), \tag{91}$$

where the fixed cost incurred with duopoly supply $(2F)$ has been ignored for now.

[210] This discussion is based on Section 3 of Armstrong and Sappington (2006). See Auriol and Laffont (1992) and Riordan (1996) for related analysis. Anton and Gertler (2004) show how a regulator can define the boundaries of local monopoly jurisdictions according to the costs reported by two local monopolists.

[211] 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 $(1-\rho)/2$.

[212] 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.

If monopoly regulation is chosen, it proceeds as in the Baron and Myerson framework of Section 2.3.1. As discussed in Section 2.3.3 (see expressions (28) and (29)), the maximum expected welfare in the regulated monopoly regime is

$$\frac{1}{2}v(c_L) + \frac{1}{2}v\big(c_H + (1-\alpha)\Delta^c\big), \tag{92}$$

abstracting from the fixed cost of monopoly supply (F) and the fixed cost of monopoly regulation (G) for now.

Regulated monopoly offers three potential advantages over unregulated duopoly 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) economies of scale in supply are preserved because there is only one industry supplier.[213] Unregulated duopoly also offers three potential advantages in this setting. First, the likelihood that the industry producer has the low marginal cost is higher under duopoly than under monopoly (unless $\rho = 1$) due to the sampling benefit of competition: if one firm fails to secure the low cost, its rival may do so. Second, the presence of a rival with correlated costs reduces the information advantage of the industry producer. This is the rent-reducing benefit of competition. Third, the cost of ongoing regulation, G, is avoided in the unregulated duopoly regime.

A formal comparison of these potential benefits of regulated monopoly and unregulated duopoly is facilitated initially by considering the case where $G = F = 0$. A comparison of expressions (91) and (92) provides three insights regarding the relative performance of regulated monopoly and unregulated duopoly.

First, as noted above, unregulated duopoly delivers a higher level of expected social welfare than does regulated monopoly when the duopolists' costs are perfectly correlated (so $\rho = 1$).[214] 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 cost. 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.[215] When demand is perfectly inelastic, price distortions do not change output levels, and therefore do not affect the

[213] In addition, when there is a social cost of public funds ($\Lambda > 0$), the firm's profit can be taxed to reduce the overall public tax burden. This benefit is not available in the case of unregulated duopoly.

[214] If $\alpha = 1$, the two regimes provide the same expected social welfare.

[215] When demand is perfectly inelastic (so $Q(p) \equiv 1$, for example), expression (91) is weakly greater than (92) whenever $\frac{1}{2}\rho \geqslant (\rho - \frac{1}{2})\alpha$, which is always the case.

firm's rent or total surplus. 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.[216]

Third, when demand is sufficiently elastic (and $\rho < 1$), regulated monopoly will generate a higher level of expected welfare than unregulated duopoly. When demand is 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.[217]

The discussion to this point has abstracted from fixed costs of supply. When the fixed cost F is sufficiently large, regulated monopoly will outperform unregulated duopoly in the simple model analyzed here because monopoly avoids the duplication of the fixed cost.[218] The analysis to this point also has assumed that both firms find it profitable to operate in the unregulated duopoly setting. If their marginal costs are highly correlated, two unregulated suppliers of a homogeneous product may not find it profitable to operate in the industry, even if fixed costs are not particularly large. This is because the firms earn no profit when their costs are identical, and so expected profit will be meager when costs are likely to be identical. Consequently, financial subsidies will be necessary to attract competing suppliers of homogeneous products when their costs are highly correlated. This situation can provide a coherent argument for assisting entry into the market.[219] The requisite subsidies will tend to be smaller when industry price competition is less intense, as it can be when the firms' products are not homogeneous, for example.

Obviously, this simple, illustrative comparison of the relative performance of regulation and competition is far from complete. A complete comparison would need to consider more carefully the policy instruments available to the regulator, for example.[220] The foregoing discussion presumes the regulator can tax consumers to finance transfer payments to the firm. In practice (as emphasized throughout Section 3), regulators are not always able to make transfer payments to the firms they regulate. Absent this ability, a regulator who wishes to ensure the monopolist never terminates its operations in the present setting can do no better than to set a single price equal to the high

[216] A similar insight is that unregulated duopoly outperforms regulated monopoly when Δ^c, the difference between the high and the low marginal cost, is 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.

[217] The convexity of $v(\cdot)$ implies that welfare in (92) is no lower than $\frac{1}{2}v(c_L)+\frac{1}{2}v(c_H)-\frac{1}{2}(1-\alpha)\Delta^c Q(c_H)$. Hence, the difference between (92) and (91) is at least $\frac{1}{2}(1-\rho)(v(c_L)-v(c_H))-[\frac{1-\alpha}{2}+\alpha(1-\rho)]\Delta^c Q(c_H)$. Since this expression is increasing in α, it is at least $\frac{1}{2}(1-\rho)(v(c_L)-v(c_H))-\frac{1}{2}\Delta^c Q(c_H)$. Therefore, welfare is higher with regulated monopoly whenever $v(c_L) \geqslant v(c_H)+\Delta^c Q(c_H)/(1-\rho)$. This inequality will hold if demand is sufficiently elastic.

[218] In contrast, if the fixed cost of regulation, G, is sufficiently large, unregulated duopoly will outperform regulated monopoly.

[219] See Armstrong and Sappington (2006) for additional discussion of the merits and drawbacks to various forms of entry assistance.

[220] A complete analysis also should consider the social cost of public funds (Λ). When Λ is large, the taxable profit generated under regulated monopoly can increase welfare substantially.

marginal cost $p = c_H$ (assuming fixed costs are zero). This policy generates expected welfare

$$v(c_H) + \frac{1}{2}\alpha \Delta^c Q(c_H). \tag{93}$$

The level of expected welfare in expression (93) is lower than the corresponding level for duopoly in expression (91).[221] Consequently, unregulated duopoly is always preferred to this restricted form of monopoly regulation (when fixed costs of supply are not significant). More generally, the relative benefits of monopoly regulation may be diminished by important practical limitations such as the possibility of regulatory capture, imperfect regulatory commitment powers, and more severe information asymmetries, for example.

Finally, in the simple static setting analyzed here, franchise bidding can secure the benefits of both regulated monopoly and unregulated duopoly, and thereby outperform both regimes. To prove this conclusion formally, return to the setting of Section 4.2. For simplicity, suppose there are two potential bidders and the two possible cost realizations are equally likely for each firm. In addition, suppose the two firms' cost realizations are independent, so that $\rho = \frac{1}{2}$. (Otherwise, the full-information outcome is possible by making use of the yardstick reporting schemes discussed in Section 4.2.2.) Consequently, a firm with the low marginal cost will be selected to serve as the monopoly supplier with probability $\frac{3}{4}$. From Proposition 13, the maximum expected welfare in the franchise bidding setting is

$$\frac{3}{4}v(c_L) + \frac{1}{4}v\big(c_H + (1-\alpha)\Delta^c\big), \tag{94}$$

where the fixed cost of supply (F) and any costs of awarding and enforcing the franchise auction have been ignored. It is apparent that expression (94) exceeds expression (92) due to the sampling benefit of competition. Expression (94) also exceeds expression (91) when $\rho = \frac{1}{2}$.[222]

In summary, franchise bidding outperforms monopoly regulation and duopoly competition in this simple setting by securing the key benefits of both regimes. Franchise bidding permits transfer payments and price regulation to pursue social goals. It also ensures the benefit of scale economies by selecting a single supplier. In addition, franchise bidding secures the sampling and rent-reducing benefits of competition. Of course, this simple model has abstracted from several of the important potential drawbacks to franchise bidding discussed in Section 4.2. These problems include the practical difficulties associated with specifying all relevant contingencies in a franchise contract and with motivating incumbent suppliers to undertake sunk investments when regulatory expropriation is a concern.

221 The difference between expressions (91) and (93) is $\frac{\rho}{2}(v(c_L) - v(c_H)) + \alpha(\frac{1}{2} - \rho)\Delta^c Q(c_H)$. From the convexity of $v(\cdot)$, this difference is at least $[\frac{\rho}{2} - (\rho - \frac{1}{2})\alpha]\Delta^c Q(c_H) \geqslant 0$.

222 The convexity of $v(\cdot)$ implies that welfare in (94) is no lower than $\frac{3}{4}v(c_L) + \frac{1}{4}v(c_H) - \frac{1}{4}(1-\alpha)\Delta^c Q(c_H)$. Consequently, the difference between (94) and (91) is at least $\frac{1}{2}(v(c_L) - v(c_H)) - \frac{1}{4}(1+\alpha)\Delta^c Q(c_H) \geqslant 0$. The final inequality arises from the convexity of $v(\cdot)$.

4.4.2. The optimal number of industry participants

To consider the optimal number of industry participants more generally, consider a different setting where the only form of industry regulation is a determination of the number of operating licenses that are awarded. It is well known that a *laissez-faire* policy toward entry will often induce too many firms to enter, and so, in principle, entry restrictions could increase welfare.[223] In practice, of course, it is an informationally demanding task to assess both the optimal number of competitors and the identity of the "best" competitors. The latter problem may be resolved in some settings by auctioning to the highest bidders a specified number of operating licenses.[224]

In some circumstances, the regulator will choose to issue fewer licenses than he would in the absence of asymmetric knowledge of operating costs. The reason for doing so is to encourage more intense bidding among potential operators. When potential operators know that a large number of licenses will be issued, they have limited incentive to bid aggressively for a license for two reasons. First, when many licenses are available, a potential supplier is relatively likely to be awarded a license even if it does not bid aggressively for a license. Second, the value of a license is diminished when many other licenses are issued because the increased competition that results when more firms operate in the industry reduces the rent that accrues to each firm. Therefore, to induce more aggressive bidding for the right to operate (and thereby secure greater payments from potential operators that can be distributed to taxpayers), a regulator may restrict the number of licenses that he issues, thereby creating a relatively concentrated industry structure.[225]

Entry policy also can affect the speed with which consumers are served. Consider, for example, a setting where firms must incur sunk costs in order to operate, and where

[223] As Mankiw and Whinston (1986) demonstrate, excess entry can arise because an individual firm does not internalize the profit reductions that its operation imposes on other firms when it decides whether to enter an industry. The authors also show that excess entry may not arise when firms supply differentiated products. Vickers (1995b) shows that excess entry may not arise when firms have different operating costs. In this case, market competition generally affords larger market shares to the least-cost suppliers (which is related to the sampling benefit of competition). The inability of firms to capture the entire consumer surplus derived from their operation in retail markets can result in an inefficiently small amount of entry. Ghosh and Morita (2004) show that less than the socially efficient level of entry may occur in a wholesale (as opposed to a retail) industry, even when all inputs and outputs are homogeneous. Upstream entry generates downstream profit, which is not internalized by the upstream entrant in the authors' model.

[224] McMillan (1994), McAfee and McMillan (1996), Cramton (1997), Milgrom (1998), Salant (2000), and Hoppe, Jehiel and Moldovanu (2006) discuss some of the complex issues that arise in designing auctions of spectrum rights. Fullerton and McAfee (1999) analyze how best to auction rights to participate in an R&D contest. They find that it is often optimal to auction licenses to two firms, who subsequently compete to innovate.

[225] This basic conclusion arises in a variety of settings, including those analyzed by Auriol and Laffont (1992), Dana and Spier (1994), and McGuire and Riordan (1995). Also see Laffont and Tirole (2000, pp. 246–250). Wilson (1979) and Anton and Yao (1989, 1992) identify a related, but distinct, drawback to allowing firms to bid for portions of a project rather than the whole project. When split awards are possible, firms can implicitly coordinate their bids and share the surplus they thereby extract from the procurer.

firms have different marginal costs of production. If a regulator were simply to authorize a single, randomly-selected firm to operate, duplication of sunk costs would be avoided and consumers could be served immediately. However, the least-cost supplier might not be chosen to operate under this form of regulated monopoly. Under a *laissez-faire* policy regarding entry, firms may be reluctant to enter the industry for fear of facing intense competition from lower-cost rivals. Under plausible conditions, there is an equilibrium in this setting in which a low-cost firm enters more quickly than does a high-cost firm. Consequently, if all potential operators have high costs, entry may be delayed. Therefore, monopoly may be preferred to unfettered competition when immediate production is highly valued.[226]

In concluding this section, we note that the discussion to this point has abstracted from the possibility of regulatory capture. This possibility can introduce a bias toward competition and away from monopoly. To see why, consider a setting where a political principal relies on advice from a (better informed) regulator to determine whether additional competition should be admitted into the regulated industry. Because increased competition typically reduces the rent a regulated firm can secure, the firm will have an incentive to persuade the regulator to recommend against allowing additional competition. To overcome this threat of regulatory capture, it can be optimal to bias policy in favor of competition by, for example, introducing additional competition even when the regulator recommends against doing so.[227]

4.5. *Integrated versus component production*

In multiproduct industries, regulators often face the task of determining which firms will supply which products. In particular, the regulators must assess the advantages and disadvantages of integrated production and component production. Under integrated production, a single firm supplies all products. Under component production, different firms supply the different products.

One potential advantage of component production is that it may admit yardstick competition which, as indicated in Section 4.1, can limit substantially the rent of regulated suppliers. One obvious potential advantage of integrated production is that it may allow technological economies of scope to be realized. Integrated production can also give rise to *informational* economies of scope in the presence of asymmetric information. To illustrate the nature of these informational economies of scope, first consider the following simple setting with independent products.

[226] See Bolton and Farrell (1990). The authors do not consider the possibility of auctioning the monopoly franchise. When franchise auctions are feasible, their use can increase the benefits of monopoly relative to oligopoly.

[227] See Laffont and Tirole (1993a) for a formal analysis of this effect. Thus, although the possibility of capture might be expected to reduce the likelihood of entry, it acts to increase the likelihood of entry once the political principal has responded appropriately to the threat. Recall the corresponding observation in Section 2.4.2.

4.5.1. Independent products

In the setting with independent products, consumer demand for each product does not depend on the prices of the other products. To illustrate most simply how informational economies of scope can arise under integrated production in this setting, suppose there are many independent products.[228] Suppose further that each product is produced with a constant marginal cost that is observed by the producer, but not by the regulator. In addition, it is common knowledge that the cost realizations are independently distributed across the products. In this setting, the full-information outcome can be closely approximated when a single firm produces all of the regulated products. To see why, suppose the single integrated firm is permitted to choose its prices for the products it supplies. Further suppose the firm is awarded (as a transfer payment) the entire consumer surplus its prices generate. For the reasons identified in Section 2.3.1, the firm will set prices equal to marginal costs under this regulatory policy.[229] Of course, the firm will enjoy significant rent under the policy. The rent is socially costly when the regulator places more weight on consumer surplus than on rent. However, the aggregate realized rent is almost independent of the firm's various cost realizations when there are many products, each produced with an independent marginal cost. Consequently, the regulator can recover this rent for consumers by imposing a lump-sum tax on the firm (almost) equal to its expected rent, thereby approximating the full-information outcome.

In this simple setting, no role for yardstick competition arises because cost realizations are not correlated. To examine the comparison between integrated and component production when yardstick effects are present, recall the two-product framework discussed in Section 2.4.3.[230] The analysis in that section derived the optimal regulatory regime under integrated production. Now consider the optimal regime under component production. First, consider the benchmark case in which the cost realizations for the two products are independently distributed. Absent cost correlation, there is no role for yardstick competition, and the optimal regulatory regime is just the single-product regime specified in Proposition 1, applied separately to the supplier of each product. It is feasible for the regulator to choose this regime under integrated production as well. However, part (iii) of Proposition 5 indicates that the regulator can secure a higher level of welfare with a different regime. Therefore, when costs are independently distributed, integrated production is optimal.

Now suppose there is some correlation between the costs of producing the two products. If the firms are risk neutral and there are no restrictions on the losses a firm can bear, the discussion in Section 4.1.2 reveals that the full-information outcome is possible with yardstick competition, and so component production is optimal, provided the two producers do not collude. In contrast, the full-information outcome will not be attainable

[228] The following discussion, found in Dana (1993), also applies naturally to the subsequent discussion about complementary products.

[229] See Loeb and Magat (1979).

[230] The following discussion is based on Dana (1993).

if the firms must receive non-negative rent for all cost reports (due to limited liability concerns, for example). However, when the correlation between the two costs is strong, the penalties required to achieve a desirable outcome are relatively small. Consequently, limits on feasible penalties will not prevent the regulator from securing a relatively favorable outcome when the firms' costs are highly correlated. In contrast, when costs are nearly independently distributed, bounds on feasible penalties will preclude the regulator from achieving a significantly higher level of welfare under yardstick regulation than he can secure by regulating each firm independently. It is therefore intuitive (and can be shown formally) that component production is preferable to integrated production only when the correlation between cost realizations is sufficiently high.[231,232] When the correlation between costs is high, part (ii) of Proposition 5 reveals that the best policy under integrated production is to treat each firm as an independent single-product monopolist. Yardstick competition can secure a higher level of expected welfare, even when there are limits on the losses that firms can be forced to bear.

The relative merits of integrated and component production can also be investigated in a franchise auction context. For instance, suppose there are two independent franchise markets, 1 and 2, and the regulator must decide whether to auction access to the two markets in separate auctions or to "bundle" the markets together in a single franchise auction. Suppose there are two potential operators, A and B, each of which can operate in one or both markets. Suppose the cost of providing the specified service in market m is c_i^m for firm i, where $m = 1, 2$ and $i = A, B$. Further, suppose there are no economies (or diseconomies) of scope in joint supply, so that firm i's cost of supplying both markets is $c_i^1 + c_i^2$. Suppose the regulator wishes to ensure production in each market, and so imposes no reserve price in the auction(s).

If the regulator awards the franchise for the two markets by means of two separate second-price auctions, he will have to pay the winner(s)

$$\max\{c_A^1, c_B^1\} + \max\{c_A^2, c_B^2\}. \tag{95}$$

[231] Ramakrishnan and Thakor (1991) provide a related analysis in a moral hazard setting. In moral hazard settings, integrated production can provide insurance to the risk averse agent, particularly when the cost realizations are not too highly correlated. Thus, as in Dana's (1993) model of adverse selection, a preference for integrated production tends to arise in moral hazard settings when the cost realizations are not too highly correlated. The reason for the superiority of integrated production is similar in the two models: the variability of the uncertainty is less pronounced under integrated production.

[232] Riordan and Sappington (1987b) provide related findings in a setting where production proceeds sequentially, and the supplier of the second input does not learn the cost of producing the second input until after production of the first input has been completed. When costs are positively correlated, integrated production increases the agent's incentive to exaggerate his first-stage cost. This is because a report of high costs in the first stage amounts to a prediction of high costs in the second stage. Since integrated production thereby makes it more costly for the regulator to induce truthful reporting of first-stage costs, the regulator prefers component production. In contrast, integrated production can reduce the agent's incentives to exaggerate first-stage costs when costs are negatively correlated. The countervailing incentives that ensue can lead the regulator to prefer integrated production when costs are negatively correlated.

If the regulator awards the two markets by means of a second-price single auction, he will have to pay the winner

$$\max\{c_A^1 + c_A^2, c_B^1 + c_B^2\},$$

which is always (weakly) less than the amount in expression (95). Therefore, the regulator will pay less when he bundles the two franchise markets in a single auction than when he conducts two separate auctions (with potentially two different winners). This conclusion reflects the rent-reducing benefit of integrated production.[233]

4.5.2. *Complementary products*

Now, suppose there is a single final product that is produced by combining two essential inputs.[234] For simplicity, suppose the inputs are perfect complements, so one unit of each input is required to produce one unit of the final product. Consumer demand for the final product is perfectly inelastic at one unit up to a known reservation price, so the regulator procures either one unit of the final product or none of the product. The cost of producing a unit of the final product is the sum of the costs of producing a unit of each of the inputs, so again there are no technological economies of scope. The cost of producing each input is the realization of an independently distributed random variable. Therefore, there is no potential for yardstick competition under component production in this setting.[235]

In this setting, the regulator again prefers integrated production to component production. To see why most simply, suppose the cost for each input can take on one of two values, c_L or c_H, where $c_L < c_H$. Also suppose the probability of obtaining a low-cost outcome is ϕ, and the costs of producing the two inputs are independently distributed. Further suppose it is optimal to supply one unit of the final product except when both inputs have a high cost.[236]

First consider integrated production, and let R_{ij} denote the rent of the integrated firm when it has cost c_i for the first input and cost c_j for the second input. Since the

[233] This discussion is based on Palfrey (1983), who shows that the ranking between integrated and component production may be reversed when there are more than two bidders. Notice that when the two markets are awarded as a bundle, inefficient production may occur because the firm with the lowest total cost is not necessarily the firm with the lowest cost in each market. For additional analysis of the optimal design of multiproduct auctions, see Armstrong (2000), for example.

[234] The following discussion is based on Baron and Besanko (1992) and Gilbert and Riordan (1995).

[235] See Jansen (1999) for an analysis of the case where the costs of the two inputs are correlated and when, as in Dana (1993), limited liability constraints bound feasible penalties. Jansen, like Dana, concludes that when the extent of correlation is high, the benefits of yardstick competition outweigh the informational economies of scope of integrated production.

[236] It is straightforward to show that if it is optimal to ensure production for all cost realizations, the regulator has no strict preference between component production and integrated production. When supply is essential, the regulator must pay the participants the sum of the two highest possible cost realizations under both industry structures. (The same is true if production is desirable only when both components are produced with low cost.)

regulator optimally terminates operations when both costs are c_H, he can limit the firm's rent to zero when it has exactly one high-cost realization, so $R_{LH} = R_{HL} = 0$. Then, as in expression (42), the incentive constraint that ensures the firm does not claim to have exactly one high-cost realization when it truly has two low-cost realizations is $R_{LL} \geqslant \Delta^c \equiv c_H - c_L$. Since the probability of having low costs for both products is ϕ^2, the regulator must allow the integrated firm an expected rent of

$$R_{INT} = \phi^2 \Delta^c.$$

Now consider component production. Suppose a firm receives the expected transfer T_i when it claims to have $c = c_i$. If one firm reports that it has low costs, then production definitely takes place since the regulator is prepared to tolerate one high-cost realization. Consequently, the expected rent of a low-cost firm is $R_L = T_L - c_L$. If a firm reports that it has a high cost, then production takes place only with probability ϕ (i.e., when the other firm reports a low cost), and so that firm's expected rent is $R_H = T_H - \phi c_H$. The regulator will afford no rent to a firm when it has high cost, so $R_H = 0$. The incentive compatibility constraint for the low-cost firm is $R_L \geqslant T_H - \phi c_L = R_H + \phi \Delta^c = \phi \Delta^c$. Therefore, the regulator must deliver an expected rent of $\phi^2 \Delta^c$ to each firm under component production, yielding a total expected industry rent of

$$R_{COMP} = 2\phi^2 \Delta^c.$$

Thus, the regulator must deliver twice as much rent under component production than he delivers under integrated production, and so integrated production is the preferred industry structure.[237]

The regulator's preference for integrated production in this setting arises because integration serves to limit a firm's incentive to exaggerate its cost. It does so by forcing the firm to internalize an externality. The regulator disciplines the suppliers in this setting by threatening to terminate their operation if total reported cost is too high. Termination reduces the profit that can be generated on both inputs. Under component production, a firm that exaggerates its cost risks only the profit it might secure from producing a single input. Each supplier ignores the potential loss in profit its own cost exaggeration may impose on the other supplier, and so is not sufficiently reticent about cost exaggeration. In contrast, under integrated production, the single supplier considers the entire loss in profit that cost exaggeration may engender, and so is more reluctant to exaggerate costs.[238]

[237] Baron and Besanko (1992) and Gilbert and Riordan (1995) show that the regulator's preference for integrated production persists in some settings where consumer demand for the final product is not perfectly inelastic. However, Da Rocha and de Frutos (1999) report that the regulator may prefer component production to integrated production when the supports of the independent cost realizations are sufficiently disparate.

[238] This result might be viewed as the informational analogue of the well-known conclusion that component production of complementary products results in higher (unregulated) prices and lower welfare than integrated production – see Cournot (1927). As such, the result for complementary products is perhaps less surprising than the corresponding result for independent products.

4.5.3. Substitute products

To identify a setting in which component production is preferred to integrated production, suppose consumers view the two products as perfect substitutes.[239] Further suppose that consumers wish to consume at most one unit of the product. The cost of producing this unit can be either low (c_L) or high (c_H). The probability of a low cost realization is ϕ, and the production costs for the two versions of the product are independently distributed. The regulator wishes to ensure supply of the product, and is considering whether to mandate integrated production (where a single firm supplies both versions of the product) or component production.

Under integrated production, the regulator must deliver transfer payment c_H to the single firm to ensure the product is supplied. Consequently, the firm secures rent Δ^c unless its cost of producing each version of the product is c_H, which occurs with probability $(1-\phi)^2$. The integrated firm's expected rent is therefore

$$R_{INT} = \left(1 - (1-\phi)^2\right)\Delta^c.$$

Under component production, the regulator can ensure the supply of the product by employing the auction mechanism described in Section 4.2 (specialized to the case of inelastic demand). Expression (90) shows that the industry rent with component production is

$$R_{COMP} = \phi(1-\phi)\Delta^c.$$

Rent is clearly lower under component production than under integrated production. Thus, the rent-reducing benefit of competition provides a strict preference for component production (i.e., for competition) over integrated production when products are close substitutes.

4.5.4. Conclusion

The simple environments considered in this section suggest two broad conclusions regarding the optimal structure of a regulated industry. First, component production will tend to be preferred to integrated production when the costs of producing inputs are highly correlated, since the yardstick competition that component production admits can limit rent effectively. Second, integrated production will tend to be preferred to component production when the components are better viewed as complements than

[239] The following discussion is closely related to the discussion in Section 4.4.1, where demand is inelastic. The differences are that here: (i) the integrated monopoly firm has two chances to obtain a low cost realization (so there is no sampling benefit of competition); (ii) the duopoly is regulated; and (iii) the duopoly firms do not know each other's cost realization.

as substitutes. In this case, integrated production can avoid what might be viewed as a double marginalization of rents that arises under component production.[240,241]

4.6. Regulating quality with competing suppliers

When a firm's service quality is verifiable, standard auction procedures for monopoly franchises can be modified to induce the delivery of high quality services. For example, the regulator can announce a rule that specifies how bids on multiple dimensions of performance (e.g., price and service quality) will be translated into a uni-dimensional score. The regulator can also announce the privileges and obligations that will be assigned to the firm that submits the winning score. For example, the winning bidder might be required to implement either the exact performance levels that he bid or the corresponding performance promised by the bidder with the second-highest score. The optimal scoring rule generally does not simply reflect customers' actual valuations of the relevant multiple performance dimensions. Different implicit valuations are employed to help account for the different costs of motivating different performance levels. These costs include the rents that potential producers can command from their superior knowledge of their ability to secure performance on multiple dimensions.[242]

The regulator's task is more difficult when a firm's performance on all relevant dimensions of service quality is not readily measured. In this case, financial rewards and penalties cannot be linked directly to the levels of delivered service quality. When quality is not verifiable, standard procedures such as competitive bidding that work well to select least-cost providers may not secure high levels of service quality. A competitive bidding procedure may award a monopoly franchise to a producer not because the producer is more able to serve customers at low cost, but because the producer's low costs reflect the limited service quality it delivers to customers. Consequently, when quality is not verifiable, consumers may be better served when the regulator engages in individual negotiations with a randomly chosen firm than when he implements a competitive bidding process.[243]

[240] See Severinov (2003) for a more detailed analysis of the effects of substitutability or complementarity on the relative merits of component and integrated production. Cost information is assumed to be uncorrelated across the two activities, so there is no potential for yardstick comparisons. The paper also discusses an industry configuration in which the regulator deals with one firm, which sub-contracts with the second firm.

[241] Iossa (1999) analyzes a model where the information asymmetries concern consumer demand rather than cost and where only one firm has private information under component production. In this framework, integrated production tends to be preferred when the products are substitutes whereas component production tends to be preferred when the products are complements.

[242] See Che (1993), Cripps and Ireland (1994) and Branco (1997) for details. Asker and Cantillon (2005) consider a setting where firms have multi-dimensional private information about their costs of supply.

[243] Manelli and Vincent (1995) derive this conclusion in a setting where potential suppliers are privately informed about the exogenous quality of their product. Their conclusion that it is optimal to assign the same probability of operation to all potential suppliers is related to the conclusion in Section 2.3.3 regarding the optimality of pooling. In Manelli and Vincent's model, incentive compatibility considerations imply that a

Unverifiable quality need not be as constraining when production by multiple suppliers is economical. In this case, if consumers can observe the level of quality delivered by each supplier (even though quality is unverifiable), market competition can help to ensure that reasonably high levels of quality and reasonably low prices arise in equilibrium.[244,245]

4.7. Conclusions

The discussion in this section has delivered two key messages. First, actual or potential competition can greatly assist a regulator in his attempts to secure a high level of consumer surplus. Competition can serve to reduce industry operating costs and reduce the rents of industry operators. Second, competition can complicate the design of regulatory policy considerably. For example, unregulated competitors may undermine pricing structures that are designed to recover fixed operating costs efficiently or to pursue distributional objectives. The presence of multiple potential operators also introduces complex considerations with regard to the design of industry structure.[246] The optimal design of regulatory policy in the presence of potential or actual competition can entail many subtleties and can require significant knowledge of the environments in which regulated and unregulated suppliers operate. An important area for future research is the design of regulatory policy when the regulator has little information about the nature and extent of competitive forces.

firm with a low-quality product, and thus low operating costs, must be selected to operate at least as often as is a firm with a high-quality product, and thus high operating costs. However, welfare is higher when high-quality products are produced. This fundamental conflict between what incentive compatibility concerns render feasible and what is optimal is resolved by a compromise in which all potential suppliers have the same probability of being selected to operate, regardless of their costs (and thus the quality of their product). Hart, Shleifer and Vishny (1997) provide the related observation that when key performance dimensions are not contractible, supply by a public enterprise may be preferable to supply by a private enterprise. The public enterprise's reduced focus on profit can lead it to supply a higher level of the costly non-contractible performance dimension (e.g., quality) than the private enterprise.

244 Because imperfect competition generally directs too few consumers to the most efficient producer, a regulator with substantial knowledge of firms' costs and consumers' preferences may prefer to set market boundaries for individual producers rather than allow market competition to determine these boundaries. [See Anton and Gertler (2004).] When the regulator's information is more limited, he may prefer to allow competitive forces to determine the customers that each firm serves. (See Wolinsky (1997).)

245 In network settings where the final quality delivered to consumers depends on the quality of all network components, producers of some network components may "free-ride" on the quality delivered by the producers of other network components. Consequently, realized quality may fall below the ideal level of quality, and a regulator may optimally allow monopoly supply of all network components in order to overcome the free-rider problem. See Auriol (1998) for further details.

246 Industry structure, in turn, can affect realized service quality. In the electric power industry, for example, recent trends toward vertical disintegration and reduced horizontal concentration have led to concerns about system reliability. See Joskow (1997, 2004), for example.

5. Vertical relationships

Regulated industries rarely take the simple form that has been assumed throughout much of the preceding discussion. Regulated industries often encompass several complementary segments that differ in their potential for competition.[247] For instance, an industry might optimally entail monopolistic supply of essential inputs (e.g., network access) but admit competitive supply of retail services. In such a setting, competitors will require access to the inputs produced in the monopolistic sector if they are to offer retail services to consumers.

Figure 27.4 illustrates two important policy issues that arise in such a setting. The first question, addressed in Section 5.1, concerns the terms on which rivals should be afforded access to the inputs supplied by the monopolist. A key consideration is how these terms should vary according to the extent of the monopolist's participation in the retail market, whether the monopolist's retail tariff is regulated, whether rivals can operate using inputs other than the input supplied by the monopolist, and whether the rivals are regulated. The second question, addressed in Section 5.2, is whether the monopolist should be permitted to operate in the potentially competitive retail market. Section 5.3 extends the discussion of access pricing to a setting where competing firms wish to purchase essential inputs from each other.

The discussion in most of this section presumes the regulator is fully informed about industry demand and cost conditions. This departure from preceding discussions reflects both the focus in the literature and the complexity of the issues raised by vertical relationships even in the presence of complete information.

5.1. One-way access pricing

Before analyzing (in Section 5.1.2) the optimal access pricing policy when the monopolist is vertically integrated, consider the simpler case where the input supplier does not operate downstream.[248] If the downstream industry is competitive in the sense that there is a negligible markup of the retail price over marginal cost, then pricing access at cost is approximately optimal. The reason is that the markup of the retail price over the total cost of providing the end-to-end service will be close to zero in this setting. In contrast, if the downstream market is not perfectly competitive, it may be optimal (if feasible) to price access below cost in order to induce lower downstream prices, which exceed marginal costs due to the imperfect competition.

[247] For an account of the theory of vertical relationships in an unregulated context, see Rey and Tirole (2007).

[248] See, for instance, Armstrong, Cowan and Vickers (1994, Section 5.2.1) and Laffont and Tirole (2000, Section 2.2.5). See Armstrong (2002, Section 2) for a more detailed account of the theory of access pricing, from which Section 5.1 is taken. See Vogelsang (2003) for a complementary review of the recent literature on access pricing in network industries. Section 5.1 abstracts from the possibility that the monopolist may try to disadvantage downstream rivals using various non-price instruments. Section 5.2 considers this possibility.

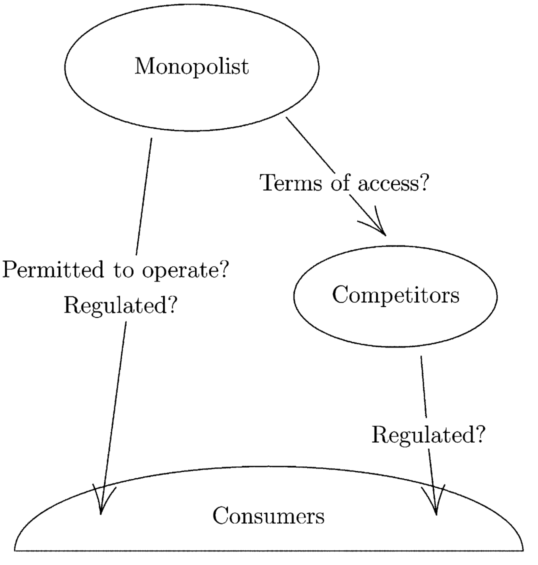

Figure 27.4. Vertical relationships.

5.1.1. The effect of distorted retail tariffs

The retail prices charged by regulated firms often depart significantly from underlying marginal costs. Section 4.3 suggested two reasons for such divergence. First, marginal-cost prices will generate negative profit for a firm that operates with increasing returns to scale. Second, regulated retail tariffs may be set to promote universal service or to redistribute income. The ensuing discussion in this section considers the impact on welfare and entry of regulated prices that diverge from cost.

The interaction between distorted tariffs and entry is illustrated most simply by abstracting from vertical issues. Therefore, suppose initially that the regulated firm's rivals do not need access to any inputs supplied by the regulated firm to provide their services. As in Section 4.3, consider a competitive fringe model in which the same service is offered by a group of rivals. Competition within the fringe means that the competitors' prices reflect their operating costs, and the fringe makes no profit.[249]

Suppose the fringe and the regulated firm offer differentiated products to final consumers. Let P and p denote the regulated firm's price and the fringe's price for their respective retail services. (Throughout this section, variables that pertain to the regulated firm will be indicated by upper-case letters. Variables that pertain to the fringe

[249] If entrants had market power, access charges should be chosen with the additional aim of controlling the retail prices of entrants. This would typically lead to a reduction in access charges. The lower input costs reduce equilibrium prices, thereby counteracting the entrants' market power. See Laffont and Tirole (2000, Section 3.3.1) for a discussion of this issue.

will be denoted by lower-case letters.) Let $V(P, p)$ be total consumer surplus when prices P and p are offered. The consumer surplus function satisfies the envelope conditions $V_P(P, p) = -X(P, p)$ and $V_p(P, p) = -x(P, p)$, where $X(\cdot,\cdot)$ and $x(\cdot,\cdot)$ are, respectively, the demand functions for the services of the regulated firm and the fringe. (Subscripts denote partial derivatives.) Assume the two services are substitutes, so $X_p(\cdot, \cdot) \equiv x_P(\cdot, \cdot) \geqslant 0$. The regulated firm has constant marginal cost C_1 and the fringe has marginal (and average) cost c. To implement the optimal output from the fringe, suppose the regulator levies a per-unit output tax t on the fringe's service. Then competition within the fringe ensures the fringe's equilibrium price is $p = c + t$. Suppose the regulated firm's price is fixed exogenously at P. Also suppose the regulator seeks to maximize total unweighted surplus (including tax revenue).[250] Total surplus in this setting is

$$W = \underbrace{V(P, c+t)}_{\text{consumer surplus}} + \underbrace{tx(P, c+t)}_{\text{tax revenue}} + \underbrace{(P - C_1)X(P, c+t)}_{\text{regulated firm's profits}}. \tag{96}$$

Maximizing W with respect to t reveals the optimal output tax for the fringe is

$$t = \sigma_d(P - C_1), \tag{97}$$

where

$$\sigma_d = \frac{X_p}{-x_p} \geqslant 0 \tag{98}$$

is a measure of the substitutability of the two retail services. In particular, σ_d measures the reduction in the demand for the regulated firm's service when the fringe supplies one additional unit of its service (and where this increase in fringe supply is caused by a corresponding reduction in its price p). If increased fringe supply comes primarily from reduced sales by the regulated incumbent, then $\sigma_d \approx 1$. If increased fringe supply largely reflects an expansion of total output with little reduction in the regulated firm's output, then $\sigma_d \approx 0$. Equation (97) implies that when sales are profitable for the regulated firm ($P > C_1$) it is optimal to raise the fringe's price above cost as well ($t > 0$). This is because profits are socially valuable, and when $P > C_1$ it is optimal to stimulate demand for the regulated firm's service in order to increase its profit. This stimulation is achieved by increasing the fringe's price. A *laissez-faire* policy towards entry ($t = 0$) would induce excessive fringe supply if the market is profitable for the regulated firm and insufficient fringe supply if the regulated firm incurs a loss in the market. Of course, if the regulated firm's price reflects its cost ($P = C_1$), then there is no need to impose an output tax on the fringe.

In expression (97), the tax t is set equal to the profit the regulated firm foregoes when fringe supply increases by a unit. This lost profit (or "opportunity cost") is the product

[250] Because there is no asymmetric information in this analysis, there is no reason to leave the regulated firm with rent. Consequently, maximization of total surplus is an appropriate objective.

of: (1) the regulated firm's unit profit $(P - C_1)$ from the sale of its product; and (2) σ_d, the reduction in the regulated firm's final sales caused by increasing fringe output by one unit. If the services are not close substitutes (σ_d is close to zero), this optimal tax should also be close to zero, and a *laissez-faire* policy towards rivals is nearly optimal. This is because a tax on the fringe's sales has little impact on the welfare associated with the sales of the regulated firm, and therefore there is little benefit from causing the fringe's price to diverge from cost.[251]

5.1.2. *Access pricing with exogenous retail prices for the monopolist*

Now return to our primary focus on vertically-related markets where the fringe requires access to inputs supplied monopolistically by the regulated firm. This section considers how best to set access prices, given the prevailing tariff for the monopolist's retail products. Ideally, a regulator would set retail prices and access prices simultaneously. (See Section 5.1.3.) However, in practice, retail tariffs often are dictated by historical, political, or social considerations, and regulators are compelled to set access prices, taking retail tariffs as given.

Suppose the monopolist supplies its retail service at constant marginal cost C_1, and supplies its access service to the fringe at constant marginal cost C_2. Let P denote the (exogenous) price for the monopolist's retail service and a denote the per-unit price paid by the fringe for access to the monopolist's input. Suppose that when it pays price a for access the fringe has the constant marginal cost $\psi(a)$ for producing a unit of its own retail service. (The cost $\psi(a)$ includes the payment a per unit of access to the monopolist.) If the fringe cannot bypass the monopolist's access service, so that exactly one unit of access is needed for each unit of its final product, then $\psi(a) \equiv a + c$, where c is the fringe's cost of converting the input into its retail product. If the fringe can substitute away from the access service, then $\psi(a)$ is a concave function of a. Note that $\psi'(a)$ is the fringe's demand for access per unit of its retail service (by Shephard's lemma). Therefore, when it supplies x units of service to consumers, the fringe's demand for access is $\psi'(a)x$. Note also that the end-to-end marginal cost of supplying one unit of the fringe's output when the price for access is a is[252]

$$\hat{c}(a) \equiv \psi(a) - (a - C_2)\psi'(a). \tag{99}$$

[251] Given the welfare function (96), it makes little difference whether the proceeds from the tax t are passed directly to the regulated firm, to the government, or into an industry fund. If the regulated firm has historically used the proceeds from a profitable activity to finance loss-making activities or to cover its fixed costs, then the firm will not face funding problems as a result of the fringe's entry if the fringe pays the tax to the firm. A more transparent mechanism would be to use a "universal service" fund to finance loss-making services. See Armstrong (2002, Section 2.1) for details. More generally, see Braeutigam (1979, 1984) and Laffont and Tirole (1993b, ch. 5) for discussions of Ramsey pricing in the presence of competition, including cases where rivals are regulated.

[252] The fringe incurs the per-unit cost $\psi(a)$, while for each unit of fringe supply the monopolist receives revenue from the fringe equal to $a\psi'(a)$ and incurs the production cost $C_2\psi'(a)$.

Whenever the fringe has some ability to substitute away from the monopolist's input, i.e., when $\psi'' \neq 0$, the access pricing policy that minimizes the fringe supply cost $\hat{c}(a)$ entails marginal-cost pricing of access: $a = C_2$. Only in this case will the fringe make the efficient input choice. In the special case where the fringe cannot substitute away from the monopolist's input ($\psi(a) \equiv a + c$), the choice of a has no impact on productive efficiency and $\hat{c}(a) \equiv C_2 + c$.

The following analysis proceeds in two stages. First, the optimal policy is derived in the case where the regulator has a full range of policy instruments with which to pursue his objectives. Second, the optimal policy is analyzed in the setting where the regulator's sole instrument is the price of access.

Regulated fringe price Suppose the regulator can control both the price of access and the fringe's retail price. When the regulator levies a per-unit output tax t on the fringe, its retail price is $p = t + \psi(a)$. Then, much as in expression (96), total welfare is

$$W = \underbrace{V\big(P, t + \psi(a)\big)}_{\text{consumer surplus}} + \underbrace{(P - C_1)X\big(P, t + \psi(a)\big)}_{\text{monopoly's profits from retail}} + \underbrace{(a - C_2)\psi'(a)x\big(P, t + \psi(a)\big)}_{\text{monopoly's profits from access}} + \underbrace{tx\big(P, t + \psi(a)\big)}_{\text{tax revenue}}. \tag{100}$$

Since $p = t + \psi(a)$, the regulator can be viewed as choosing p and a rather than t and a. In this case, expression (100) simplifies to

$$W = V(P, p) + (P - C_1)X(P, p) + \big(p - \hat{c}(a)\big)x(P, p), \tag{101}$$

where $\hat{c}(\cdot)$ is given in expression (99). Since a does not affect any other aspect of welfare in expression (101), a should be chosen to minimize $\hat{c}(\cdot)$. In particular, whenever the fringe can substitute away from the monopolist's input ($\psi''(a) \neq 0$), it is optimal to set $a = C_2$. Also, maximizing expression (101) with respect to $p = t + \psi(a)$ yields formula (97) for the output tax t. In sum, the optimal policy involves

$$a = C_2; \qquad t = \sigma_d(P - C_1). \tag{102}$$

When the regulator can utilize an output tax to control the fringe's supply, access should be priced at cost, and the fringe's output tax should be equal to the monopolist's opportunity cost of fringe supply, as given in (97). In contrast, if the fringe had access to the monopolist's input at cost but did not have to pay an output tax, then (just as in Section 5.1.1) there would be excess fringe supply if $P > C_1$ and insufficient fringe supply if $P < C_1$. There would, however, be no productive inefficiency under this policy, and the fringe's service would be supplied at minimum cost. As before, if the monopolist's retail price is equal to its cost ($P = C_1$), there is no need to regulate the fringe and a policy to allow entrants to purchase access at marginal cost is optimal.

Provided the regulator has enough policy instruments to pursue all relevant objectives, there is no need to sacrifice productive efficiency even when the monopolist's retail price differs from its cost. An output tax on rivals can be used to counteract incumbent retail tariffs that diverge from cost, while pricing access at marginal cost can ensure productive efficiency in rival supply.

Unregulated fringe price Now consider the optimal policy when the access price is the sole instrument available to the regulator.[253] In this case, $t = 0$ in (100), and welfare when the price for access is a is

$$W = \underbrace{V\big(P, \psi(a)\big)}_{\text{consumer surplus}} + \underbrace{(P - C_1)X\big(P, \psi(a)\big)}_{\text{monopoly's profits from retail}} + \underbrace{(a - C_2)\psi'(a)x\big(P, \psi(a)\big)}_{\text{monopoly's profits from access}}. \quad (103)$$

Notice that in this setting, the only way the regulator can ensure a high price for the fringe's output (perhaps because he wishes to protect the monopolist's socially valuable profit as outlined in Section 5.1.1) is to set a high price for access. The high access price typically will cause some productive inefficiency in fringe supply.

Maximizing expression (103) with respect to a reveals that the optimal price of access is

$$a = C_2 + \sigma(P - C_1), \quad (104)$$

where

$$\sigma = \frac{X_p \psi'(a)}{-z_a} \geqslant 0 \quad (105)$$

and $z(P, a) \equiv \psi'(a)x(P, \psi(a))$ is the fringe's demand for access. The "displacement ratio" σ measures the reduction in demand for the monopolist's retail service caused by supplying the marginal unit of access to the fringe.[254] Therefore, expression (104) states that the price of access should be equal to the cost of access, C_2, plus the monopolist's foregone profit (opportunity cost) caused by supplying a unit of access to its rivals. This rule is sometimes known as the "efficient component pricing rule" (or ECPR).[255]

Considerable information is required to calculate the displacement ratio σ. In practice, a regulator may have difficulty estimating this ratio accurately. The social costs of the estimation errors can be mitigated by bringing the monopolist's retail price P closer to its cost, C_1. From expression (104), doing so lessens the dependence of the access price on σ. As before, if $P = C_1$, it is optimal to price access at cost.

In the special case where consumer demands for the two retail services are approximately independent (so $X_p(\cdot,\cdot) \approx 0$), formula (104) states that the access price

[253] This discussion is based on Armstrong, Doyle and Vickers (1996).

[254] The access charge a must fall by $1/z_a$ to expand the fringe demand for access by one unit. This reduction in a causes X to decline by $X_p\psi'/z_a$.

[255] This rule appears to have been proposed first in Willig (1979). See Baumol (1983), Baumol and Sidak (1994a, 1994b), Baumol, Ordover and Willig (1997), Sidak and Spulber (1997), and Armstrong (2002, Section 2.3.1) for further discussions of the ECPR.

should involve no mark-up over the cost of providing access, even if $P \neq C_1$. In other cases, however, the optimal price for access is not equal to the associated cost. Consequently, there is productive inefficiency whenever there is some scope for substitution ($\psi''(a) \neq 0$). The inefficiency arises because a single instrument, the access price, is forced to perform two functions – to protect the monopolist's socially valuable profit, and to induce fringe operators to employ an efficient mix of inputs – and the regulator must compromise between these two objectives.

This analysis is simplified in the special case where the fringe cannot substitute away from the monopolist's input, so $\psi(a) \equiv c + a$. In this case, expression (104) becomes

$$a = C_2 + \sigma_d(P - C_1), \tag{106}$$

where σ_d is the demand substitution parameter given in expression (98). Expression (106) states that the optimal access price is the sum of the cost of providing access and the optimal output tax given in expression (97). Thus, an alternative way to implement the optimum in this case would be to price access at cost (C_2) and simultaneously levy a tax on the output of rivals, as in expression (102). When exactly one unit of access is needed to produce one unit of fringe output (so there is no scope for productive inefficiency), this output tax could also be levied on the input. In this case, the regulator's lack of policy instruments is not constraining. More generally, however, the regulator can achieve a strictly higher level of welfare if he can choose both an output tax and an access price.

5.1.3. Ramsey pricing

Now consider the optimal simultaneous choice of the monopolist's retail and access prices.[256] The optimal policy will again depend on whether rivals can substitute away from the input, and, if they can, on the range of policy instruments available to the regulator.

Regulated fringe price Suppose first that the regulator can set the price for the monopolist's retail product P, impose a per-unit output tax t on the fringe, and set price a for the input. Also suppose the proceeds of the output tax are used to contribute to the financing of the monopolist's fixed costs. As before, the price of the fringe's product is equal to the perceived marginal cost, so $p = t + \psi(a)$. The regulator can again be considered to choose p rather than t. Letting $\lambda \geqslant 0$ be the Lagrange multiplier on the monopolist's profit constraint, the regulator's problem is to choose P, p and a to maximize

$$W = V(P, p) + (1 + \lambda)\big[(P - C_1)X(P, p) + \big(p - \hat{c}(a)\big)x(P, p)\big]. \tag{107}$$

[256] See Laffont and Tirole (1994) and Armstrong, Doyle and Vickers (1996).

For any retail prices, P and p, the access price a affects only the production cost of the fringe. Consequently, a should be chosen to minimize the cost of providing the fringe's service, $\hat{c}(a)$. As before, whenever the fringe can substitute away from the input ($\psi'' \neq 0$), the optimal policy is to price access at cost (so $a = C_2$).[257] The two retail prices, P and p, are then chosen to maximize total surplus subject to the monopolist's profit constraint.

Unregulated fringe price Now suppose the regulator cannot impose an output tax on the fringe. In this case, $p = \psi(a)$, and the access price takes on the dual role of attempting to ensure the fringe employs an efficient input mix and influencing the fringe retail price in a desirable way. Following the same logic that underlies expressions (103) and (107), welfare in this setting can be written as

$$W = V\big(P, \psi(a)\big) + (1+\lambda)\big[(P - C_1)X\big(P, \psi(a)\big) + (a - C_2)\psi'(a)x\big(P, \psi(a)\big)\big]. \tag{108}$$

The first-order condition for maximizing expression (108) with respect to a is

$$a = \underbrace{C_2 + \sigma(P - C_1)}_{\text{ECPR price}} + \left[\frac{\lambda}{1+\lambda}\right]\frac{a}{\eta_z}, \tag{109}$$

where σ is given in expression (105), $\eta_z = -az_a/z > 0$ is the own-price elasticity of demand for access, and P is the Ramsey price for the monopolist's retail service. Expression (109) states that the optimal access price is given by the ECPR expression (104), which would be optimal if the monopolist's retail price were exogenously fixed at P, plus a Ramsey markup that is inversely related to the elasticity of fringe demand for access. This Ramsey markup reflects the benefits (in terms of a reduction in P) caused by increasing the revenue generated by selling access to the fringe. One can show that the Ramsey pricing policy entails $P > C_1$ and $a > C_2$, and so access is priced above marginal cost. Thus, a degree of productive inefficiency arises whenever the fringe can substitute away from the monopolist's input. As in Section 5.1.2, when the access price is called upon to perform multiple tasks, a compromise is inevitable.

5.1.4. Unregulated retail prices

Now consider how best to price access when the access price is the regulator's only instrument. Suppose the monopolist can set its preferred retail price P.[258] In addition,

[257] This is just an instance of the general principle that productive efficiency is optimally induced when the policy maker has a sufficient number of control instruments at his disposal. See Diamond and Mirrlees (1971).
[258] This discussion is adapted from Laffont and Tirole (1994, Section 7) and Armstrong and Vickers (1998). For other analyses of access pricing with an unregulated downstream sector, see Economides and White (1995), Lewis and Sappington (1999), Lapuerta and Tye (1999), and Sappington (2005a), for example.

suppose the fringe's price is unregulated, and the regulator cannot impose an output tax on the fringe. As before, if the regulator sets the access price a, the fringe's price is $p = \psi(a)$. Given a, the monopolist will set its retail price P to maximize its total profit

$$\Pi = (P - C_1)X\big(P, \psi(a)\big) + (a - C_2)\psi'(a)x\big(P, \psi(a)\big).$$

Let $\bar{P}(a)$ denote the monopolist's profit-maximizing retail price given access price a. The firm generally will set a higher retail price when the price of access is higher (so $\bar{P}'(a) > 0$). This is the case because the more profit the monopolist anticipates from selling access to its rivals, the less aggressively the firm will compete with rivals at the retail level. Social welfare is given by expression (103), where the monopolist's price P is given by $\bar{P}(a)$. The welfare-maximizing access price in this setting satisfies

$$a = \underbrace{C_2 + \sigma\big(\bar{P}(a) - C_1\big)}_{\text{ECPR price}} - \frac{X\bar{P}'}{-z_a}, \tag{110}$$

where σ is given in expression (105). Equation (110) reveals that the optimal access price in this setting is below the level in the ECPR expression (104), which characterizes the optimal access price in the setting where the monopolist's retail price was fixed at $\bar{P}(a)$. The lower access charge is optimal here because a reduction in a causes the retail price P to fall towards cost, which increases welfare.[259]

In general, it is difficult to determine whether the price of access, a, is optimally set above, at, or below the marginal cost of providing access, C_2. However, there are three special cases where the price of access should equal cost: (1) when the fringe has no ability to substitute away from the input ($\psi(a) \equiv a + c$) and the demand functions $X(\cdot,\cdot)$ and $x(\cdot,\cdot)$ are linear; (2) when the monopolist and fringe operate in separate retail markets, with no cross-price effects[260]; and (3) when the fringe and the monopolist offer the same homogeneous product, i.e., when the retail market is potentially perfectly competitive. To understand the third (and most interesting) of these cases, consider a setting where consumers purchase from the supplier that offers the lowest retail price. If the price of access is a, the fringe will supply consumers whenever the monopolist offers a retail price greater than the fringe's cost, $\psi(a)$. Therefore, given a, the monopolist has two options. First, it can preclude fringe entry by setting a retail price just below $\psi(a)$. Doing so ensures it a profit of $\psi(a) - C_1$ per unit of retail output. Second, the monopolist can choose not to supply the retail market itself, and simply sell access to the fringe. By following this strategy, the monopolist secures profit $(a - C_2)\psi'(a)$ per unit of retail output. The monopolist will choose to accommodate entry if and only if the

[259] By contrast, it was optimal to raise the access charge above (104) in the Ramsey problem – see expression (109) above. This is because an increase in the access charge allowed the incumbent's retail price to fall, since the access service then financed more of the regulated firm's costs.

[260] The profit-maximizing retail price $\bar{P}$ does not depend on a in this case. Also, since $\sigma = 0$, expression (110) implies that marginal-cost pricing of access is optimal.

latter profit margin exceeds the former, i.e., if

$$C_1 \geqslant \hat{c}(a),$$

where $\hat{c}(\cdot)$ is given in expression (99). Therefore, the monopolist will allow entry by the fringe if and only if retail supply by the fringe is less costly than supply by the monopolist (when the access price is a). Perhaps surprisingly, then, the choice of access price does not bias competition for or against the fringe in this setting. The reason is that, although the monopolist's actual cost of supplying consumers is C_1, its opportunity cost of supplying consumers is $\hat{c}(a)$, since it then foregoes the profits from selling access to the fringe.[261] However, the choice of access price has a direct effect on the equilibrium retail price, which is $\psi(a)$ regardless of which firm supplies consumers. Consequently, when the fringe firms are the more efficient suppliers (i.e., $\psi(C_2) \leqslant C_1$), it is optimal to provide access to the fringe at cost. Doing so will ensure a retail price equal to the minimum cost of production and supply by the least-cost supplier.[262]

More generally, when the monopolist has some market power in the retail market, the optimal access price will equal marginal cost only in knife-edge cases. Clear-cut results are difficult to obtain in this framework because the access price is called upon to perform three tasks. It serves: (i) to control the market power of the monopolist (a lower value of a induces a lower value for the monopolist's retail price P); (ii) to protect the monopolist's socially valuable profit (as discussed in Section 5.1.1); and (iii) to pursue productive efficiency for the fringe (which requires $a = C_2$) whenever there is a possibility for substituting away from the input. In general, tasks (i) and (iii) argue for an access price no higher than cost. (When $a = C_2$ the monopolist will choose $P > C_1$. Setting $a < C_2$ will reduce its retail price towards cost. Task (iii) will mitigate, but not reverse, this incentive.) However, unless a is chosen to be so low that $P < C_1$, task (ii) will give the regulator an incentive to raise a above cost – see expression (104). These counteracting forces preclude unambiguous conclusions regarding the relative magnitudes of the optimal price of access and the cost of providing access in unregulated retail markets.

5.1.5. Discussion

Setting access prices equal to the cost of providing access offers two primary advantages. First, at least in settings where the monopolist's costs are readily estimated, this

[261] Sappington (2005a) presents a model where, given the price of access, an entrant decides whether to purchase the input from the monopolist before the two firms compete to supply consumers. The entrant's "make-or-buy" decision is shown not to depend on the price of access, as the entrant chooses to buy the input from the monopolist whenever this is the efficient choice. Therefore, the access price does not affect productive efficiency, but does affect equilibrium retail prices.

[262] If industry costs are lower when the monopolist serves the market even when the fringe can purchase access at cost (i.e., if $\psi(C_2) > C_1$), then it is optimal to subsidize access (to be precise, to set the access charge to satisfy $\psi(a) = C_1$) so that competition forces the monopolist to price its retail service at its own cost.

policy is relatively simple to implement. In particular, no information about consumer demand or rivals' characteristics is needed to calculate these prices (at least in the simple models presented above).[263] Of course, it can be difficult to measure the monopolist's costs accurately in practice, especially when the firm produces multiple products and makes durable investments.[264] Second, pricing access at cost can help to ensure that rivals adopt cost-minimizing production policies. If access is priced above cost, an entrant may construct its own network rather than purchase network services from the regulated firm, even though the latter policy entails lower social cost.[265]

In simple terms, cost-based access prices are appropriate when access prices do not need to perform the role of correcting for distortions in the monopolist's retail tariff. There are three main settings in which such a task may not be necessary:

1. If the monopolist's retail tariff reflects its marginal costs, no corrective measures are needed. In particular, access prices should reflect the marginal costs of providing access. Thus, a complete rebalancing of the monopolist's tariff (so retail prices reflect costs) can greatly simplify input pricing policies, and allow access prices to focus on the task of ensuring productive efficiency.
2. If there are distortions in the regulated tariff, but corrections to this are made by using another regulatory instrument (such as output taxes levied on rivals), then access prices should reflect costs.
3. When the monopolist operates in a vigorously competitive retail market and is free to set its own retail tariff, pricing access at cost can be optimal.

In settings other than these, pricing access at cost generally is not optimal.

5.2. *Vertical structure*

The second important policy issue is whether to allow the monopoly supplier of a regulated input to integrate downstream to supply a final product to consumers in com-

[263] See Hausman and Sidak (2005) for a more general treatment of this issue.

[264] Sidak and Spulber (1997), Hausman (1997), and Hausman and Sidak (1999, 2005), among others, discuss how the cost of capital and the irreversibility of capital investment affect the cost of providing access. They note that the cost of capital tends to increase in competitive settings, since an incumbent's capital investment may be stranded in a competitive environment. Hausman also emphasizes the asymmetric advantage an entrant may be afforded when it can purchase access at a cost that does not fully reflect the risk an incumbent supplier faces. While the incumbent must invest before all the relevant uncertainty is resolved, the entrant often can wait to decide whether to make or buy the input until after the uncertainty is resolved. Also see Laffont and Tirole (2000, Section 4.4).

[265] A bias in the opposite direction has been alleged when regulators use forms of forward-looking cost-based access pricing. In the United States, the Federal Communications Commission has decided that the major incumbent suppliers of local exchange services must unbundle network elements and make these elements available to competing suppliers at prices that reflect the incumbent's total element long-run incremental cost (TELRIC). Mandy (2002) and Mandy and Sharkey (2003), among others, explore the calculation of TELRIC prices. Hausman (1997), Kahn, Tardiff and Weisman (1999), and Weisman (2002), among others, criticize TELRIC pricing of network elements, arguing that TELRIC prices provide little incentive for competitors to build their own networks and that, in practice, regulators do not have the information required to calculate appropriate TELRIC prices.

petition with other suppliers.[266] Downstream integration by a monopoly input supplier can alter industry performance in two main ways. First, it can influence directly the welfare generated in the retail market by changing the composition of, and the nature of competition in, the retail market. Second, downstream integration can affect the incentives of the monopolist, and thereby influence indirectly the welfare generated in both the upstream and downstream industries.

First consider the effects of altering the composition of the retail sector. If retail competition is imperfect, retail supply by the input monopolist can enhance competition, thereby reducing price and increasing both output and welfare in the retail market.[267] The welfare increase can be particularly pronounced if the monopolist can supply the retail service at lower cost than other retailers.[268] Furthermore, downstream production by the input monopolist can deter some potential suppliers from entering the industry and thereby avoid duplicative fixed costs of production.[269]

Now consider how the opportunity to operate downstream can affect the incentives of the input monopolist. When it competes directly in the retail market, the monopolist generally will anticipate greater profit from its retail operations as the costs of its rivals increase. Therefore, the integrated firm may seek to increase the costs of its retail rivals. It can do this in at least two ways. First, if the regulator is uncertain about the monopolist's cost of supplying the input, the monopolist may seek to raise the costs of downstream rivals by exaggerating its input cost. If the monopolist can convince the regulator that its upstream production costs are high, the regulator may raise the price of the input, thereby increasing the operating costs of downstream competitors.[270] By increasing the incentive of the monopolist to exaggerate its access costs in this manner, vertical integration can complicate the regulator's critical control problem.[271]

Second, the integrated firm may be able to raise its rivals' costs by degrading the quality of the input it supplies, by delaying access to its input, or by imposing burdensome purchasing requirements on downstream producers, for example.[272] The regulator can affect the monopolist's incentive to raise the costs of its downstream rivals through the

[266] Section 3.5.2 discusses the merits of allowing a regulated supplier to diversify into horizontally related markets.

[267] See Hinton et al. (1998) and Weisman and Williams (2001) for assessments of this effect in the United States telecommunications industry.

[268] See Lee and Hamilton (1999).

[269] See Vickers (1995a).

[270] Bourreau and Dogan (2001) consider a dynamic model in which an incumbent vertically-integrated supplier may wish to set an access charge that is unduly low. The low access charge induces retail competitors to employ the incumbent's old technology rather than invest in a more modern, superior technology.

[271] Vickers (1995a) analyzes this effect in detail. Lee and Hamilton (1999) extend Vickers' analysis to allow the regulator to condition his decision about whether to allow integration on the monopolist's reported costs.

[272] Economides (1998) examines a setting in which the incentives for raising rivals' costs in this manner are particularly pronounced. Also see Reiffen, Schumann and Ward (2000), Laffont and Tirole (2000, Section 4.5), Mandy (2000), Beard, Kaserman and Mayo (2001), Crandall and Sidak (2002), Bustos and Galetovic (2003), Sappington and Weisman (2005), and Crew, Kleindorfer and Sumpter (2005).

regulated price for access. When the monopolist enjoys a substantial profit margin on each unit of access it sells to downstream producers, the monopolist will sacrifice considerable upstream profit if it raises the costs of downstream rivals and thereby reduces their demand for access. Therefore, the regulator may reduce any prevailing incentive to degrade quality by raising the price of access.[273] A complete assessment of optimal regulatory policy in this regard awaits further research.

The input monopolist's participation in the retail market can complicate the design of many simple, practical regulatory policies, including price cap regulation. To understand why, recall from Section 3.1.3 that price cap regulation often constrains the average level of the regulated firm's prices. An aggregate restriction on overall price levels can admit a substantial increase in the price of one service (e.g., the input sold to downstream competitors), as long as this increase is accompanied by a substantial decrease in the price of another service (e.g., the monopolist's retail price). Consequently, price cap regulation that applies to all of the prices set by an integrated monopolist could allow the firm to exercise a price squeeze. An integrated monopolist exercises a price squeeze when the margin between its retail price and its access price is not sufficient to allow an equally efficient entrant to operate profitably. As discussed in Section 3.1.3, additional restrictions on the pricing flexibility of vertically integrated firms that operate under price cap regulation often are warranted to prevent price squeezes that force more efficient competitors from the downstream market.[274]

In summary, downstream integration by a monopoly supplier of an essential input generally entails both benefits and costs. Either the benefits or the costs can predominate, depending upon the nature of downstream competition, the relevant information asymmetries, and the regulator's policy instruments. Appropriate policy, therefore, will generally vary according to the setting in which it is implemented.

5.3. *Two-way access pricing*

Different issues can arise when two established firms need to buy inputs from each other. Such a need can arise, for example, when competing communications networks require mutual interconnection to allow customers on one network to communicate with customers on the other network.[275] Even though competition for customers may be sufficiently vigorous to limit the need for explicit regulation of retail tariffs, regulation may still be needed to ensure interconnection agreements that are in the public interest.

[273] Thus, one advantage of the ECPR policy discussed in Section 5.1.2 (which can involve a significant markup of the access charge above cost) is that the firm's incentive to degrade quality is lessened, relative to a cost-based policy. See Weisman (1995, 1998), Reiffen (1998), Sibley and Weisman (1998), Beard, Kaserman and Mayo (2001), and Sand (2004) for related analyses.

[274] See Laffont and Tirole (1996) and Bouckaert and Verboven (2004).

[275] Interconnection is one aspect of the general issue of compatibility of services between rival firms. See Farrell and Klemperer (2007) for a survey which includes a discussion of this issue.

Suppose for simplicity there are two symmetric networks, A and B, and consumers wish to subscribe to one of these networks (but not both).[276] If the two firms set the same retail tariff they will attract an equal number of subscribers. In this case, suppose a subscriber on network A, say, makes half her telephone calls to subscribers on network A and half to subscribers on network B. To complete these latter calls, A will need to arrange for B to deliver the calls destined for B's subscribers that originate on A's network. Similarly, B will need to arrange for A to deliver calls to A's subscribers initiated by B's subscribers.

Consider the following timing. Suppose the network operators first negotiate their mutual access prices and then compete non-cooperatively in the retail market for subscribers, taking as given the negotiated access prices. In a symmetric setting, the negotiated prices for access will be reciprocal, so that the payment A makes to B for each of A's calls that B delivers is the same as the corresponding payment from B to A. Moreover, the two firms will have congruent preferences regarding the ideal reciprocal access price.[277] Of course, a regulator will want to know if the firms might employ interconnection arrangements to distort retail competition, and, if so, whether the negotiated access prices will exceed or fall below the access prices that are ideal from the social perspective.[278]

The answers to these questions turn out to be subtle, and to depend in part on the kinds of retail tariffs firms employ. Consider the following three types of tariff.

Linear pricing First suppose the two network operators employ linear pricing, so that they charge a single per-minute price for calls with no additional fixed (e.g., monthly) charge. In this case, firms will be tempted to choose high payments for access in order to relax competition for subscribers. The mutual benefit of a high reciprocal access price is apparent. Because firm A's customers make a fraction of their calls to customers on firm B's network, a high access charge paid to B will increase A's effective cost of providing calls to its subscribers. In equilibrium, this inflated cost induces a high equilibrium retail price. Therefore, in this setting, firms may try to negotiate a high reciprocal access price in order to implement high retail prices and thereby increase their joint profits. (Of course, in a symmetric equilibrium, firms receive access revenue from their rival exactly equal to the access payments they deliver to the rival. Consequently, the only effect of high access payments on profit is the beneficial effect of higher induced retail prices.) Therefore, an active role for regulation may persist in this setting to ensure firms do not agree to set access payments that are unduly high.

[276] The following discussion is based on Armstrong (1998), Laffont, Rey and Tirole (1998a), and Carter and Wright (1999). See Laffont and Tirole (2000, ch. 5) and Armstrong (2002, Section 4.2) for more comprehensive reviews of the literature on two-way access pricing.

[277] Carter and Wright (2003) and Armstrong (2004) demonstrate that such congruence is not ensured when firms are asymmetric.

[278] In benchmark models where network and call externalities are unimportant, for example, the socially optimal price for access is the marginal cost of providing access, because that access price induces a unit call price equal to the cost of making a call.

Two-part pricing Now consider the (more realistic) case where firms compete for subscribers using two-part tariffs (that consist of a per-minute price and a monthly fixed charge, for example). As with linear pricing, the access price affects each firm's perceived marginal cost of providing calls. Consequently, a high access price will induce a high per-minute price in equilibrium, which, in turn, will generate high profits from supplying calls. However, firms now have an additional instrument with which to compete for subscribers – the monthly fixed charge. Since high access payments ensure a high (per-subscriber) profit from providing calls, firms will compete vigorously to attract additional subscribers to their network. They will do so by setting a low fixed charge. In simple models of subscriber demand for networks (e.g., a Hotelling market for subscribers), equilibrium profit turns out to be independent of the established access price.[279] Consequently, firms have no strict preference among access prices, and so are likely to be amenable to setting socially desirable prices for access.

Price discrimination Finally, consider the case where firms can set a different price for a call depending on whether the call is made to a subscriber on the same network (an "on-net" call) or to a subscriber on the rival network (an "off-net" call).[280] (Firms continue to set a fixed monthly charge as well.) Such discriminatory tariffs can introduce important network size effects. To illustrate, suppose firms charge a higher price for off-net calls than for on-net calls. This pricing structure will induce subscribers to prefer the network with the larger number of subscribers, all else equal. This preference reflects the fact that a subscriber on the larger network can make more calls at a relatively low price. This preference gives rise to a positive network size effect, as larger firms are better able to attract subscribers than smaller firms. In contrast, suppose the price of an off-net call is lower than the price of an on-net call. A negative network size effect arises in this case, as subscribers prefer to subscribe to the smaller network, all else equal. When two established firms vie for subscribers (with no scope for market entry or exit), competition is less intense (in the sense that equilibrium prices are higher) in settings with negative network size effects than in settings with positive network size effects. Competition is less intense in the former case because a unilateral price reduction does not attract many new subscribers, since subscribers prefer not to join the larger network.

Firms can determine whether positive or negative network size effects arise through the access tariffs they set. When the firms establish a reciprocal access price in excess of the marginal cost of delivering a call, the equilibrium price for an off-net call will exceed the equilibrium price for an on-net call and a positive network size effect will arise. This effect will intensify competition for subscribers. In contrast, if the firms

[279] Dessein (2003) and Hahn (2004) show that this "profit neutrality" result extends to settings where subscribers have heterogeneous demands for calls and where firms can offer different tariffs to different consumer types. Dessein shows that the profit neutrality result does not extend to cases where the total number of subscribers is affected by the prevailing retail tariffs.

[280] Laffont, Rey and Tirole (1998b) and Gans and King (2001) analyze this possibility.

set the access price below marginal cost, the equilibrium price for an on-net call will exceed the equilibrium for an off-net call. The resulting negative network size effect will induce relatively weak competition for subscribers, and thereby increase the profit of established competitors. Therefore, in this setting, the firms will wish to price access below the socially desirable access price. The regulator's task in this setting, then, will be to ensure the firms do not negotiate access payments that are unduly low.[281,282]

In sum, the developing literature on two-way access pricing suggests a number of continuing roles for regulation, even when competition for subscribers is sufficiently vigorous to limit the need for explicit retail tariff regulation.

5.4. Conclusions

The discussion in this section has reviewed some of many subtleties that arise when a regulated firm operates at multiple stages of the production process. The discussion emphasized the standard theme that a regulator generally is better able to achieve social goals the more instruments he has at his disposal. In particular, a well-informed regulator generally can secure greater industry surplus when he can set access prices and retail prices simultaneously and when he can control the activities of all industry competitors. The discussion also emphasized the fact that even after substantial competition develops among facilities-based network operators, regulatory oversight of the interconnection agreements between these operators may be warranted.

The discussion in this section has followed the literature in focusing on settings in which the regulator is fully informed about the regulated industry. Further research is warranted on the design of regulatory policy in vertically-integrated industries where regulators are less omniscient. Additional directions for future research are suggested in Section 6.

6. Summary and conclusions

This chapter has reviewed recent theoretical studies of the design of regulatory policy, focussing on studies in which the regulated firm has better information about its environment than does the regulator. The regulator's task in such settings often is to try to

[281] "Bill-and-keep" is a common policy that implements access charges below cost. Under a bill-and-keep policy, each network agrees to complete the calls initiated on other networks without charge. While such a policy can reduce the intensity of competition, it may have a more benign effect, which is to reduce the price for making calls so that call externalities – the (often) beneficial effect of a call on the recipient of a call – are internalized. [See Hermalin and Katz (2004) and Jeon, Laffont and Tirole (2004), for example.] The presence of significant call externalities can render bill-and-keep a desirable policy. [See DeGraba (2003, 2004) and Berger (2005), for example.]

[282] Laffont, Rey and Tirole (1998b) argue that when network-based price discrimination is practiced, high access charges (and hence high off-net call charges) by incumbents can be used as an instrument to deter entry.

induce the firm to employ its superior information in the broader social interest. One central message of this chapter is that this regulatory task can be a difficult and subtle one. The regulator's ability to induce the firm to use its privileged information to pursue social goals depends upon a variety of factors, including the nature of the firm's private information, the environment in which the firm operates, the regulator's policy instruments, and his commitment powers.

Recall from Section 2, for example, that despite having limited knowledge of consumer demand, a regulator may be able to secure the ideal outcome for consumers when the regulated firm operates with decreasing returns to scale. In contrast, a regulator generally is unable to secure the ideal outcome for consumers when the regulated firm has privileged knowledge of its cost structure. However, even in this setting, a regulator with strong commitment powers typically can ensure that consumers and the firm both gain as the firm's costs decline. The regulator can do so by providing rent to the firm that admits to having lower costs. But when a regulator cannot make long-term commitments about how he will employ privileged information revealed by the firm, the regulator may be unable to induce the firm to employ its superior information to achieve Pareto gains. Thus, the nature of the firm's superior knowledge, the firm's operating technology, the regulator's policy instruments, and his commitment powers are all of substantial importance in the design of regulatory policy.

The fact that information, technology, instruments, and institutions all matter in the design of regulatory policy implies that the best regulatory policy typically will vary across industries, across countries, and over time. Thus, despite our focus in this chapter on generic principles that apply in a broad array of settings, institutional details must be considered carefully when designing regulatory policy for a specific institutional setting. Future research that transforms the general principles reviewed above to concrete regulatory policies in particular settings will be of substantial value.

Another central message of this chapter is that options constitute important policy instruments for the regulator. It is through the careful structuring of options that the regulator can induce the regulated firm to employ its privileged information to further social goals. As noted above, the options generally must be designed to cede rent to the regulated firm when it reveals that it has the superior ability required to deliver greater benefits to consumers. Consequently, it is seldom costless for the regulator to induce the regulated firm to employ its privileged information in the social interest. However, the benefits of providing discretion to the regulated firm via carefully-structured options generally outweigh the associated costs, and so such discretion typically is a component of optimal regulatory policy in the presence of asymmetric information.

This chapter has reviewed two distinct strands of the literature. Section 2 reviewed studies of the optimal design of regulatory policy in Bayesian settings. Section 3 reviewed non-Bayesian analyses of simple, practical regulatory policies and policies that have certain desirable properties in specified settings. Bayesian analyses of the optimal design of regulatory policy typically entail the structuring of options for the regulated firm. As noted above, in such analyses, the regulator employs his limited knowledge of the regulatory environment to construct a set of options, and then permits the firm to choose one of the specified options. In contrast, non-Bayesian analyses often consider the implementation of a single regulatory policy that does not present the firm with an explicit choice among options.[283] One interpretation of the non-Bayesian approach may be that regulatory plans that encompass options are "complicated", and therefore prohibitively costly to implement.[284] A second interpretation might be that the regulator has no information about the regulatory environment that he can employ to structure options for the firm. To assess the validity of this interpretation, future research might analyze the limit of optimal Bayesian regulatory policies as the regulator's knowledge of the regulatory environment becomes less and less precise. It would be interesting to determine whether any of the policies reviewed in Section 3 emerge as the limit of optimal regulatory policies in such an analysis.

Future research might also analyze additional ways to harness the power of competition to complement regulatory policy. As emphasized in Section 4, even though competition can complicate the design and implementation of regulatory policy, it can also provide pronounced benefits for consumers. The best manner in which to capture these benefits without sacrificing unduly the benefits that regulation can provide merits additional consideration, both in general and in specific institutional settings. The analysis in this chapter has focused on the substantial benefits that competition can deliver in static settings, where products and production technologies are immutable. In dynamic settings, competition may deliver better products and superior production techniques, in addition to limiting the rents of incumbent suppliers. Reasonable, if not optimal, policies to promote and harness these potential benefits of competition merit additional research, particularly in settings where the regulator's information about key elements of the regulated industry is severely limited.

In addition to examining how competition can best complement regulatory policy, future research might analyze the conditions under which competition can replace regulatory oversight. Broad conclusions regarding the general merits of deregulation and specific findings regarding the merits of deregulation in particular institutional settings would both be valuable. Most of the analyses reviewed in this chapter have taken as given the fact that a regulator will dictate the prices that a monopoly provider can charge. Two related questions warrant further study. First, how can a regulator determine when sufficient (actual or potential) competition has developed in an industry so that ongoing price regulation is no longer in the social interest? Second, when direct price regulation is no longer warranted, are other forms of regulatory oversight and control useful? For example, might ongoing monitoring of industry prices, service quality, and the state of competition usefully supplement standard antitrust policy immediately following industry deregulation?[285]

[283] Of course, the regulator provides the firm with meaningful options whenever he offers the firm some discretion over its prices, as in Section 3.1.

[284] Ideally, the costs of complexity should be modeled explicitly, and the costs of more complicated regulatory plans should be weighed against their potential benefits.

[285] Armstrong and Sappington (2006) and Gerardin and Sidak (2006), among others, discuss the interaction between regulatory and antitrust policy.

In closing, we emphasize the importance of empirical work as a complement to both the theoretical work reviewed in this chapter and future theoretical work on the design of regulatory policy.[286] Theoretical research typically models the interplay among conflicting economic forces, and specifies conditions under which one force outweighs another force. Often, though, theoretical analysis cannot predict unambiguously which forces will prevail in practice. Carefully structured empirical research can determine which forces prevailed under particular circumstances, and can thereby provide useful insight about the forces that are likely to prevail in similar circumstances. Thus, despite our focus on theoretical work in this chapter, it is theoretical work and empirical work together that ultimately will provide the most useful guidance to policy makers and the greatest insight regarding the design of regulatory policy.

Acknowledgements

We are grateful to Carli Coetzee, Simon Cowan, Ernesto Dal Bó, Jos Jansen, Paul Joskow, Rob Porter, Ray Rees, Michael Riordan, Jean Tirole, and Ingo Vogelsang for helpful comments.

References

Acton, J., Vogelsang, I. (1989). "Introduction to the symposium on price cap regulation". RAND Journal of Economics 20 (3), 369–372.

Anton, J., Gertler, P. (1988). "External markets and regulation". Journal of Public Economics 37 (2), 243–260.

Anton, J., Gertler, P. (2004). "Regulation, local monopolies and spatial competition". Journal of Regulatory Economics 25 (2), 115–142.

Anton, J., Yao, D. (1987). "Second sourcing and the experience curve: Price competition in defense procurement". RAND Journal of Economics 18 (1), 57–76.

Anton, J., Yao, D. (1989). "Split awards, procurement and innovation". RAND Journal of Economics 20 (4), 538–552.

Anton, J., Yao, D. (1992). "Coordination in split award auctions". Quarterly Journal of Economics 97 (2), 681–707.

Anton, J., Vander Weide, J., Vettas, N. (2002). "Entry auctions and strategic behavior under cross-market price constraints". International Journal of Industrial Organization 20 (5), 611–629.

Armstrong, M. (1998). "Network interconnection in telecommunications". Economic Journal 108 (448), 545–564.

Armstrong, M. (1999). "Optimal regulation with unknown demand and cost functions". Journal of Economic Theory 84 (2), 196–215.

Armstrong, M. (2000). "Optimal multi-object auctions". Review of Economic Studies 67 (3), 455–481.

Armstrong, M. (2001). "Access pricing, bypass and universal service". American Economic Review 91 (2), 297–301.

[286] Sappington (2002) provides a review of recent empirical work that examines the effects of incentive regulation in the telecommunications industry. Also see Kridel, Sappington and Weisman (1996).

Armstrong, M. (2002). "The theory of access pricing and interconnection". In: Cave, M., Majumdar, S., Vogelsang, I. (Eds.), Handbook of Telecommunications Economics, vol. I. North-Holland, Amsterdam.
Armstrong, M. (2004). "Network interconnection with asymmetric networks and heterogeneous calling patterns". Information Economics and Policy 16 (3), 375–390.
Armstrong, M., Rochet, J.-C. (1999). "Multi-dimensional screening: A user's guide". European Economic Review 43 (4), 959–979.
Armstrong, M., Sappington, D. (2004). "Toward a synthesis of models of regulatory policy design with limited information". Journal of Regulatory Economics 26 (1), 5–21.
Armstrong, M., Sappington, D. (2006). "Regulation, competition and liberalization". Journal of Economic Literature 44 (2), 325–366.
Armstrong, M., Vickers, J. (1991). "Welfare effects of price discrimination by a regulated monopolist". RAND Journal of Economics 22 (4), 571–580.
Armstrong, M., Vickers, J. (1993). "Price discrimination, competition and regulation". Journal of Industrial Economics 41 (4), 335–360.
Armstrong, M., Vickers, J. (1998). "The access pricing problem with deregulation: A note". Journal of Industrial Economics 46 (1), 115–121.
Armstrong, M., Vickers, J. (2000). "Multiproduct price regulation under asymmetric information". Journal of Industrial Economics 48 (2), 137–160.
Armstrong, M., Cowan, S., Vickers, J. (1994). Regulatory Reform: Economic Analysis and British Experience. MIT Press, Cambridge, MA.
Armstrong, M., Cowan, S., Vickers, J. (1995). "Nonlinear pricing and price cap regulation". Journal of Public Economics 58 (1), 33–55.
Armstrong, M., Doyle, C., Vickers, J. (1996). "The access pricing problem: A synthesis". Journal of Industrial Economics 44 (2), 131–150.
Armstrong, M., Rees, R., Vickers, J. (1995). "Optimal regulatory lag under price cap regulation". Revista Espanola De Economia 12, 93–116.
Asker, J., Cantillon, E. (2005). "Optimal procurement when both price and quality matter". Mimeo.
Auriol, E. (1998). "Deregulation and quality". International Journal of Industrial Organization 16 (2), 169–194.
Auriol, E., Laffont, J.-J. (1992). "Regulation by duopoly". Journal of Economics and Management Strategy 1 (3), 507–533.
Averch, H., Johnson, L. (1962). "Behavior of the firm under regulatory constraint". American Economic Review 52 (5), 1053–1069.
Bailey, E., Coleman, R. (1971). "The effect of lagged regulation in an Averch–Johnson model". Bell Journal of Economics 2 (1), 278–292.
Baron, D. (1985). "Noncooperative regulation of a nonlocalized externality". RAND Journal of Economics 16 (4), 553–568.
Baron, D. (1988). "Regulation and legislative choice". RAND Journal of Economics 29 (3), 457–477.
Baron, D. (1989). "Design of regulatory mechanisms and institutions". In: Schmalensee, R., Willig, R. (Eds.), Handbook of Industrial Organization, vol. II. North-Holland, Amsterdam.
Baron, D., Besanko, D. (1984a). "Regulation and information in a continuing relationship". Information Economics and Policy 1 (4), 267–302.
Baron, D., Besanko, D. (1984b). "Regulation, asymmetric information, and auditing". RAND Journal of Economics 15 (4), 447–470.
Baron, D., Besanko, D. (1987a). "Commitment and fairness in a dynamic regulatory relationship". Review of Economic Studies 54 (3), 413–436.
Baron, D., Besanko, D. (1987b). "Monitoring, moral hazard, asymmetric information, and risk sharing in procurement contracting". RAND Journal of Economics 18 (4), 509–532.
Baron, D., Besanko, D. (1992). "Information, control, and organizational structure". Journal of Economics and Management Strategy 1 (2), 237–275.
Baron, D., Myerson, R. (1982). "Regulating a monopolist with unknown costs". Econometrica 50 (4), 911–930.

Baseman, K. (1981). "Open entry and cross-subsidization in regulated markets". In: Fromm, G. (Ed.), Studies in Public Regulation. MIT Press, Cambridge, MA.
Baumol, W. (1983). "Some subtle issues in railroad regulation". International Journal of Transport Economics 10 (1–2), 341–355.
Baumol, W., Klevorick, A. (1970). "Input choices and rate-of-return regulation: An overview of the discussion". Bell Journal of Economics 1 (2), 162–190.
Baumol, W., Sidak, G. (1994a). "The pricing of inputs sold to competitors". Yale Journal on Regulation 11 (1), 171–202.
Baumol, W., Sidak, G. (1994b). Toward Competition in Local Telephony. MIT Press, Cambridge, MA.
Baumol, W., Bailey, E., Willig, R. (1977). "Weak invisible hand theorems on the sustainability of prices in a multiproduct monopoly". American Economic Review 67 (3), 350–365.
Baumol, W., Ordover, J., Willig, R. (1997). "Parity pricing and its critics: A necessary condition for efficiency in the provision of bottleneck services to competitors". Yale Journal on Regulation 14 (1), 145–164.
Baumol, W., Panzar, J., Willig, R. (1982). Contestable Markets and the Theory of Industry Structure. Harcourt Brace Jovanovich, Inc., New York.
Beard, R., Kaserman, D., Mayo, J. (2001). "Regulation, vertical integration, and sabotage". Journal of Industrial Economics 46 (3), 319–333.
Berg, S., Lynch, J. (1992). "The measurement and encouragement of telephone service quality". Telecommunications Policy 16 (3), 210–224.
Berger, U. (2005). "Bill-and-keep vs. cost-based access pricing revisited". Economics Letters 86 (1), 107–112.
Bernstein, J., Sappington, D. (1999). "Setting the X factor in price cap regulation plans". Journal of Regulatory Economics 16 (1), 5–25.
Besanko, D. (1985). "On the use of revenue requirements regulation under imperfect information". In: Crew, M. (Ed.), Analyzing the Impact of Regulatory Change in Public Utilities. Lexington Books, Lexington, MA.
Besanko, D., Spulber, D. (1992). "Sequential-equilibrium investment by regulated firms". RAND Journal of Economics 23 (2), 153–170.
Besanko, D., Donnenfeld, S., White, L. (1987). "Monopoly and quality distortion: Effects and remedies". Quarterly Journal of Economics 102 (4), 743–767.
Besanko, D., Donnenfeld, S., White, L. (1988). "The multiproduct firm, quality choice, and regulation". Journal of Industrial Economics 36 (4), 411–429.
Bester, H., Strausz, R. (2001). "Contracting with imperfect commitment and the revelation principle: The single agent case". Econometrica 69 (4), 1077–1098.
Biais, B., Perotti, E. (2002). "Machiavellian privatization". American Economic Review 92 (1), 240–258.
Biglaiser, G., Ma, C.-T.A. (1995). "Regulating a dominant firm: Unknown demand and industry structure". RAND Journal of Economics 26 (1), 1–19.
Biglaiser, G., Ma, C.-T.A. (1999). "Investment incentives of a regulated dominant firm". Journal of Regulatory Economics 16 (3), 215–236.
Biglaiser, G., Riordan, M. (2000). "Dynamics of price regulation". RAND Journal of Economics 31 (4), 744–767.
Blackmon, G. (1994). Incentive Regulation and the Regulation of Incentives. Kluwer Academic Publishers, Norwell, MA.
Boiteux, M. (1956). "Sur la gestion des monopoles publics astreints à l'équilibre budgétaire". Econometrica 24 (1), 22–40.
Bolton, P., Dewatripont, M. (2005). Contract Theory. MIT Press, Cambridge, MA.
Bolton, P., Farrell, J. (1990). "Decentralization, duplication, and delay". Journal of Political Economy 98 (4), 803–826.
Bouckaert, J., Verboven, F. (2004). "Price squeezes in a regulatory environment". Journal of Regulatory Economics 26 (3), 321–351.
Bourreau, M., Dogan, P. (2001). "Regulation and innovation in the telecommunications industry". Telecommunications Policy 25 (3), 167–184.

Bower, A. (1993). "Procurement policy and contracting efficiency". International Economic Review 34 (4), 873–901.
Bradley, I., Price, C. (1988). "The economic regulation of private industries by price constraints". Journal of Industrial Economics 37 (1), 99–106.
Braeutigam, R. (1979). "Optimal pricing with intermodal competition". American Economic Review 69 (1), 38–49.
Braeutigam, R. (1984). "Socially optimal pricing with rivalry and economies of scale". RAND Journal of Economics 15 (1), 127–134.
Braeutigam, R. (1989). "Optimal policies for natural monopolies". In: Schmalensee, R., Willig, R. (Eds.), Handbook of Industrial Organization, vol. II. North-Holland, Amsterdam.
Braeutigam, R. (1993). "A regulatory bargain for diversified enterprises". International Journal of Industrial Organization 11 (1), 1–20.
Braeutigam, R., Panzar, J. (1989). "Diversification incentives under 'price-based' and 'cost-based' regulation". RAND Journal of Economics 20 (3), 373–391.
Braeutigam, R., Panzar, J. (1993). "Effects of the change from rate-of-return regulation to price cap regulation". American Economic Review 83 (2), 191–198.
Branco, F. (1997). "The design of multidimensional auctions". RAND Journal of Economics 28 (1), 63–81.
Brennan, T. (1989). "Regulating by 'capping prices' ". Journal of Regulatory Economics 1 (2), 133–147.
Brennan, T. (1990). "Cross-subsidization and cost misallocation by regulated monopolists". Journal of Regulatory Economics 2 (1), 37–52.
Brennan, T., Palmer, K. (1994). "Comparing the costs and benefits of diversification by regulated firms". Journal of Regulatory Economics 6 (1), 115–136.
Bustos, A., Galetovic, A. (2003). "Vertical integration and sabotage in regulated industries". Mimeo.
Cabral, L., Riordan, M. (1989). "Incentives for cost reduction under price cap regulation". Journal of Regulatory Economics 1 (2), 93–102.
Caillaud, B. (1990). "Regulation, competition, and asymmetric information". Journal of Economic Theory 52 (1), 87–110.
Caillaud, B., Guesnerie, R., Rey, P. (1992). "Noisy observation in adverse selection models". Review of Economic Studies 59 (3), 595–615.
Caillaud, B., Guesnerie, R., Rey, P., Tirole, J. (1988). "Government intervention in production and incentives theory: A review of recent contributions". RAND Journal of Economics 19 (1), 1–26.
Carter, M., Wright, J. (1999). "Interconnection in network industries". Review of Industrial Organization 14 (1), 1–25.
Carter, M., Wright, J. (2003). "Asymmetric network interconnection". Review of Industrial Organization 22 (1), 27–46.
Celentani, M., Ganuza, J.-J. (2002). "Corruption and competition in procurement". European Economic Review 46 (7), 1273–1303.
Chang, Y.-M., Warren, J. (1997). "Allocative efficiency and diversification under price-cap regulation". Information Economics and Policy 9 (1), 3–17.
Che, Y.-K. (1993). "Design competition through multidimensional auctions". RAND Journal of Economics 24 (4), 668–680.
Che, Y.-K. (1995). "Revolving doors and the optimal tolerance for agency collusion". RAND Journal of Economics 26 (3), 378–397.
Che, Y.-K., Gale, I. (1998). "Standard auctions with financially constrained bidders". Review of Economic Studies 65 (1), 1–22.
Che, Y.-K., Gale, I. (2000). "The optimal mechanism for selling to budget-constrained consumers". Journal of Economic Theory 92 (2), 198–233.
Che, Y.-K., Gale, I. (2003). "Optimal design of research contests". American Economic Review 93 (3), 64–671.
Chiappori, P., Macho, I., Rey, P., Salanié, B. (1994). "Repeated moral hazard: The role of memory, commitment and the access to credit markets". European Economic Review 38 (8), 1527–1553.

Chu, L.Y., Sappington, D. (2007). "Simple cost-sharing contracts". American Economic Review 97 (1), 419–428.
Clemenz, G. (1991). "Optimal price cap regulation". Journal of Industrial Economics 39 (4), 391–408.
Cooper, R. (1984). "On the allocative distortions in problems of self-selection". RAND Journal of Economics 15 (4), 568–577.
Cournot, A. (1927). Researches into the Mathematical Principles of the Theory of Wealth. Macmillan, New York.
Cowan, S. (1997a). "Price-cap regulation and inefficiency in relative pricing". Journal of Regulatory Economics 12 (1), 53–70.
Cowan, S. (1997b). "Tight average revenue regulation can be worse than no regulation". Journal of Industrial Economics 45 (1), 75–88.
Cowan, S. (2004). "Optimal risk allocation for regulated monopolies and consumers". Journal of Public Economics 88 (1), 285–303.
Crampes, C., Hollander, A. (1995). "Duopoly and quality standards". European Economic Review 39 (1), 71–82.
Cramton, P. (1997). "The FCC spectrum auctions: An early assessment". Journal of Economics and Management Strategy 6 (3), 431–495.
Crandall, R., Sidak, J.G. (2002). "Is structural separation of incumbent local exchange carriers necessary for competition?". Yale Journal on Regulation 19 (2), 335–411.
Crew, M., Crocker, K. (1991). "Diversification and regulated monopoly". In: Crew, M. (Ed.), Competition and the Regulation of Utilities. Kluwer Academic Publishers, Boston, MA.
Crew, M., Kleindorfer, P. (2002). "Regulatory economics: Twenty years of progress?". Journal of Regulatory Economics 21 (1), 5–22.
Crew, M., Kleindorfer, P., Sumpter, J. (2005). "Bringing competition to telecommunications by divesting the RBOCs". In: Crew, M., Spiegel, M. (Eds.), Obtaining the Best from Regulation and Competition. Kluwer Academic Publishers, Norwell, MA.
Cripps, M., Ireland, N. (1994). "The design of auctions and tenders with quality thresholds: The symmetric case". Economic Journal 104 (423), 316–326.
Crémer, J., Khalil, F., Rochet, J.-C. (1998a). "Contracts and productive information gathering". Games and Economic Behavior 23 (2), 174–193.
Crémer, J., Khalil, F., Rochet, J.-C. (1998b). "Strategic information gathering before a contract is offered". Journal of Economic Theory 81 (1), 163–200.
Crémer, J., McLean, R. (1985). "Optimal selling strategies under uncertainty for a discriminating monopolist when demands are interdependent". Econometrica 53 (2), 345–361.
Curien, N., Jullien, B., Rey, P. (1998). "Pricing regulation under bypass competition". RAND Journal of Economics 29 (2), 259–279.
Currier, K. (2005). "Strategic firm behavior under average-revenue-lagged regulation". Journal of Regulatory Economics 27 (1), 67–80.
Da Rocha, J., de Frutos, M.A. (1999). "A note on the optimal structure of production". Journal of Economic Theory 89 (2), 234–246.
Dalen, D.M. (1997). "Regulation of quality and the ratchet effect: Does unverifiability hurt the regulator?". Journal of Regulatory Economics 11 (2), 139–155.
Dalen, D.M. (1998). "Yardstick competition and investment incentives". Journal of Economics and Management Strategy 7 (1), 105–126.
Dana, J. (1993). "The organization and scope of agents: Regulating multiproduct industries". Journal of Economic Theory 59 (2), 288–310.
Dana, J., Spier, K. (1994). "Designing a private industry: Government auctions with endogenous market structure". Journal of Public Economics 53 (1), 127–147.
Dasgupta, S., Spulber, D. (1989/1990). "Managing procurement auctions". Information Economics and Policy 4 (1), 5–29.
De Fraja, G., Iozzi, A. (2004). "Bigger and better: A dynamic regulatory mechanism for optimum quality". Mimeo.

DeGraba, P. (2003). "Efficient intercarrier compensation on competing networks when customers share the value of a call". Journal of Economics and Management Strategy 12 (2), 207–230.
DeGraba, P. (2004). "Reconciling the off-net pricing principle with efficient network utilization". Information Economics and Policy 16 (3), 475–494.
Demougin, D., Garvie, D. (1991). "Contractual design with correlated information under limited liability". RAND Journal of Economics 22 (4), 477–489.
Demsetz, H. (1968). "Why regulate utilities?". Journal of Law and Economics 11 (1), 55–65.
Demski, J., Sappington, D. (1984). "Optimal incentive contracts with multiple agents". Journal of Economic Theory 33 (1), 152–171.
Demski, J., Sappington, D. (1987). "Hierarchical regulatory control". RAND Journal of Economics 18 (3), 369–383.
Demski, J., Sappington, D., Spiller, P. (1987). "Managing supplier switching". RAND Journal of Economics 18 (1), 77–97.
Demski, J., Sappington, D., Spiller, P. (1988). "Incentive schemes with multiple agents and bankruptcy constraints". Journal of Economic Theory 44 (1), 156–167.
Dessein, W. (2003). "Network competition in nonlinear pricing". RAND Journal of Economics 34 (4), 593–611.
Diamond, P., Mirrlees, J. (1971). "Optimal taxation and public production. I. Production efficiency". American Economic Review 61 (1), 8–27.
Dobbs, I. (2004). "Intertemporal price cap regulation under uncertainty". Economic Journal 114 (495), 421–440.
Eckel, C., Lutz, N. (2003). "Introduction: What role can experiments play in research on regulation?". Journal of Regulatory Economics 23 (2), 103–108.
Economides, N. (1998). "The incentive for non-price discrimination by an input monopolist". International Journal of Industrial Organization 16 (3), 271–284.
Economides, N., White, L. (1995). "Access and interconnection pricing: How efficient is the efficient component pricing rule?". The Antitrust Bulletin 40 (3), 557–579.
Einhorn, M. (1987). "Optimality and sustainability: Regulation and intermodal competition in telecommunications". RAND Journal of Economics 18 (4), 550–563.
Encinosa, W., Sappington, D. (1995). "Toward a benchmark for optimal prudency policy". Journal of Regulatory Economics 7 (2), 111–131.
Farrell, J., Klemperer, P. (2007). "Coordination and lock-in: Competition with switching costs and network effects". In: Armstrong, M., Porter, R. (Eds.), Handbook of Industrial Organization, vol. 3. North-Holland, Amsterdam (this volume).
Faure-Grimaud, A., Martimort, D. (2003). "Regulatory inertia". RAND Journal of Economics 34 (3), 413–437.
Finsinger, J., Vogelsang, I. (1981). "Alternative institutional frameworks for price incentive mechanisms". Kyklos 34 (3), 338–404.
Finsinger, J., Vogelsang, I. (1982). "Performance indices for public enterprises". In: Jones, L. (Ed.), Public Enterprise in Less Developed Countries. Cambridge Univ. Press, Cambridge.
Flores, D. (2005). "Price cap regulation in the Mexican telephone industry". Information Economics and Policy 17 (2), 231–246.
Foreman, D. (1995). "Pricing incentives under price-cap regulation". Information Economics and Policy 7 (4), 331–351.
Fraser, R. (1995). "The relationship between the costs and the prices of a multi-product monopoly: The role of price-cap regulation". Journal of Regulatory Economics 8 (1), 23–32.
Freixas, X., Guesnerie, R., Tirole, J. (1985). "Planning under incomplete information and the ratchet effect". Review of Economic Studies 52 (2), 173–191.
Fudenberg, D., Tirole, J. (1990). "Moral hazard and renegotiation in agency contracts". Econometrica 58 (6), 1279–1320.
Fudenberg, D., Tirole, J. (1991). Game Theory. MIT Press, Cambridge, MA.

Fullerton, R., McAfee, P. (1999). “Auctioning entry into tournaments”. Journal of Political Economy 107 (3), 573–605.

Gans, J., King, S. (2001). “Using ‘bill and keep’ interconnect agreements to soften network competition”. Economics Letters 71 (3), 413–420.

Gary-Bobo, R., Spiegel, Y. (2006). “Optimal state-contingent regulation under limited liability”. RAND Journal of Economics 37 (2), 431–448.

Gasmi, F., Ivaldi, M., Laffont, J.-J. (1994). “Rent extraction and incentives for efficiency in recent regulatory proposals”. Journal of Regulatory Economics 6 (2), 151–176.

Gasmi, F., Laffont, J.-J., Sharkey, W. (1999). “Empirical evaluation of regulatory regimes in local telecommunications markets”. Journal of Economics and Management Strategy 8 (1), 61–93.

Gerardin, D., Sidak, J.G. (2006). “European and American approaches to antitrust remedies and the institutional design of regulation in telecommunications”. In: Majumdar, S., Vogelsang, I., Cave, M. (Eds.), In: Handbook of Telecommunications Economics, vol. II. North-Holland, Amsterdam.

Ghosh, A., Morita, H. (2004). “Free entry and social efficiency under vertical oligopoly”. Mimeo.

Gilbert, R., Newbery, D. (1994). “The dynamic efficiency of regulatory constitutions”. RAND Journal of Economics 25 (4), 538–554.

Gilbert, R., Riordan, M. (1995). “Regulating complementary products: A comparative institutional approach”. RAND Journal of Economics 26 (2), 243–256.

Glover, J. (1994). “A simpler method that stops agents from cheating”. Journal of Economic Theory 62 (1), 221–229.

Greenwald, B. (1984). “Rate base selection and the structure of regulation”. RAND Journal of Economics 15 (1), 85–95.

Guesnerie, R., Laffont, J.-J. (1984). “A complete solution to a class of principle-agent problems with an application to the control of a self-managed firm”. Journal of Public Economics 25 (3), 329–369.

Guthrie, G. (2006). “Regulating infrastructure: The impact on risk and investment”. Journal of Economic Literature 44 (4), 925–972.

Hagerman, J. (1990). “Regulation by price adjustment”. RAND Journal of Economics 21 (1), 72–82.

Hahn, J.-H. (2004). “Network competition and interconnection with heterogeneous subscribers”. International Journal of Industrial Organization 22 (1), 611–631.

Harris, M., Townsend, R. (1981). “Resource allocation under asymmetric information”. Econometrica 49 (1), 33–64.

Hart, O., Shleifer, A., Vishny, R. (1997). “The proper scope of government: Theory and an application to prisons”. Quarterly Journal of Economics 113 (4), 1127–1161.

Hausman, J. (1997). “Valuing the effect of regulation on new services in telecommunications”. In: Brookings Papers on Economic Activity: Microeconomics, pp. 1–38.

Hausman, J., Sidak, J.G. (1999). “A consumer-welfare approach to the mandatory unbundling of telecommunications networks”. Yale Law Journal 109 (3), 417–505.

Hausman, J., Sidak, J.G. (2005). “Telecommunications regulation: Current approaches with the end in sight”. Mimeo.

Hermalin, B., Katz, M. (2004). “Sender or receiver: Who should pay to exchange an electronic message?”. RAND Journal of Economics 35 (3), 423–447.

Hillman, J., Braeutigam, R. (1989). Price Level Regulation for Diversified Public Utilities. Kluwer Academic Publishers, Boston, MA.

Hinton, P., Zona, J.D., Schmalensee, R., Taylor, W. (1998). “An analysis of the welfare effects of long-distance market entry by an integrated access and long-distance provider”. Journal of Regulatory Economics 13 (2), 183–196.

Holmstrom, B., Milgrom, P. (1987). “Aggregation and linearity in the provision of intertemporal incentives”. Econometrica 55 (2), 303–328.

Hoppe, H., Jehiel, P., Moldovanu, B. (2006). “License auctions and market structure”. Journal of Economics and Management Strategy 15 (2), 371–396.

Inderst, R. (2002). “Contract design and bargaining power”. Economics Letters 74 (2), 171–176.

Iossa, E. (1999). “Informative externalities and pricing in regulated multiproduct industries”. Journal of Industrial Economics 47 (2), 195–220.
Iossa, E., Stroffolini, F. (2002). “Price cap regulation and information acquisition”. International Journal of Industrial Organization 20 (7), 1013–1036.
Iossa, E., Stroffolini, F. (2005). “Price cap regulation, revenue sharing and information acquisition”. Information Economics and Policy 17 (2), 217–230.
Isaac, R.M. (1991). “Price cap regulation: A case study of some pitfalls of implementation”. Journal of Regulatory Economics 3 (2), 193–210.
Jansen, J. (1999). “Regulating complementary input supply: Cost correlation and limited liability”. Mimeo. Berlin.
Jeon, D.-S., Laffont, J.-J., Tirole, J. (2004). “On the receiver pays principle”. RAND Journal of Economics 35 (1), 85–110.
Joskow, P. (1997). “Restructuring, competition and regulatory reform in the U.S. electricity sector”. Journal of Economic Perspectives 11 (3), 119–138.
Joskow, P. (2004). “The difficult transition to competitive electricity markets in the U.S.”. In: Griffin, J., Puller, S. (Eds.), Electricity Deregulation: Where to Go from Here? University of Chicago Press, Chicago.
Joskow, P. (2005). “Incentive regulation in theory and practice: Electricity and distribution networks”. Mimeo.
Jullien, B. (2000). “Participation constraints in adverse selection models”. Journal of Economic Theory 93 (1), 1–47.
Kahn, A., Tardiff, T., Weisman, D. (1999). “The 1996 Telecommunications Act at three years: An economic evaluation of its implementation by the FCC”. Information Economics and Policy 11 (4), 319–365.
Kang, J., Weisman, D., Zhang, M. (2000). “Do consumers benefit from tighter price cap regulation?”. Economics Letters 67 (1), 113–119.
Kerschbamer, R. (1994). “Destroying the “pretending” equilibria in the Demski–Sappington–Spiller model”. Journal of Economic Theory 62 (1), 230–237.
Khalil, F. (1997). “Auditing without commitment”. RAND Journal of Economics 28 (4), 629–640.
Kim, J. (1997). “Inefficiency of subgame optimal entry regulation”. RAND Journal of Economics 28 (1), 25–36.
Kim, J.-C., Jung, C.-Y. (1995). “Regulating a multi-product monopolist”. Journal of Regulatory Economics 8 (3), 299–307.
Kjerstad, E., Vagstad, S. (2000). “Procurement auctions with entry of bidders”. International Journal of Industrial Organization 18 (8), 1243–1257.
Kofman, F., Lawarrée, J. (1993). “Collusion in hierarchical agency”. Econometrica 61 (3), 629–656.
Kolbe, L., Tye, W. (1991). “The Duquesne opinion: How much ‘hope’ is there for investors in regulated firms”. Yale Journal of Regulation 8 (1), 113–157.
Kridel, D., Sappington, D., Weisman, D. (1996). “The effects of incentive regulation in the telecommunications industry: A survey”. Journal of Regulatory Economics 9 (3), 269–306.
Kwoka, J. (1991). “Productivity and price caps in telecommunications”. In: Einhorn, M. (Ed.), Price Caps and Incentive Regulation in Telecommunications. Kluwer Academic Publishers, Boston.
Kwoka, J. (1993). “Implementing price caps in telecommunications”. Journal of Policy Analysis and Management 12 (4), 726–752.
Laffont, J.-J. (2005). Regulation and Development. Cambridge Univ. Press, Cambridge, UK.
Laffont, J.-J., Martimort, D. (1997). “Collusion under asymmetric information”. Econometrica 65 (4), 875–912.
Laffont, J.-J., Martimort, D. (1999). “Separation of regulators against collusive behavior”. RAND Journal of Economics 30 (2), 232–262.
Laffont, J.-J., Martimort, D. (2002). The Theory of Incentives: The Principal–Agent Model. Princeton Univ. Press, Princeton, NJ.
Laffont, J.-J., Rochet, J.-C. (1998). “Regulation of a risk averse firm”. Games and Economic Behavior 25 (2), 149–173.
Laffont, J.-J., Tirole, J. (1986). “Using cost observation to regulate firms”. Journal of Political Economy 94 (3), 614–641.

Laffont, J.-J., Tirole, J. (1987). "Auctioning incentive contracts". Journal of Political Economy 95 (5), 921–937.

Laffont, J.-J., Tirole, J. (1988a). "The dynamics of incentive contracts". Econometrica 56 (5), 1153–1176.

Laffont, J.-J., Tirole, J. (1988b). "Repeated auctions of incentive contracts, investment, and bidding parity with an application to takeovers". RAND Journal of Economics 19 (4), 516–537.

Laffont, J.-J., Tirole, J. (1990a). "Adverse selection and renegotiation in procurement". Review of Economic Studies 57 (4), 597–626.

Laffont, J.-J., Tirole, J. (1990b). "Optimal bypass and cream skimming". American Economic Review 80 (4), 1041–1051.

Laffont, J.-J., Tirole, J. (1990c). "The politics of government decision-making: Regulatory institutions". Journal of Law, Economics, and Organization 6 (1), 1–32.

Laffont, J.-J., Tirole, J. (1991a). "Auction design and favoritism". International Journal of Industrial Organization 9 (1), 9–42.

Laffont, J.-J., Tirole, J. (1991b). "The politics of government decision-making: A theory of regulatory capture". Quarterly Journal of Economics 106 (4), 1089–1127.

Laffont, J.-J., Tirole, J. (1991c). "Privatization and incentives". Journal of Law, Economics, and Organization 7 (3), 84–105.

Laffont, J.-J., Tirole, J. (1993a). "Cartelization by regulation". Journal of Regulatory Economics 5 (2), 111–130.

Laffont, J.-J., Tirole, J. (1993b). A Theory of Incentives in Procurement and Regulation. MIT Press, Cambridge, MA.

Laffont, J.-J., Tirole, J. (1994). "Access pricing and competition". European Economic Review 38 (9), 1673–1710.

Laffont, J.-J., Tirole, J. (1996). "Creating competition through interconnection: Theory and practice". Journal of Regulatory Economics 10 (3), 227–256.

Laffont, J.-J., Tirole, J. (2000). Competition in Telecommunications. MIT Press, Cambridge, MA.

Laffont, J.-J., Rey, P., Tirole, J. (1998a). "Network competition. I. Overview and nondiscriminatory pricing". RAND Journal of Economics 29 (1), 1–37.

Laffont, J.-J., Rey, P., Tirole, J. (1998b). "Network competition. II. Price discrimination". RAND Journal of Economics 29 (1), 38–56.

Lapuerta, C., Tye, W. (1999). "Promoting effective competition through interconnection policy". Telecommunications Policy 23 (2), 129–145.

Law, P. (1995). "Tighter average revenue regulation can reduce consumer welfare". Journal of Industrial Economics 42 (4), 399–404.

Law, P. (1997). "Welfare effects of pricing in anticipation of Laspeyres price cap regulation: An example". Bulletin of Economic Research 49 (1), 17–27.

Lee, S.-H. (1997a). "A note on regulating a multiproduct monopolist". Journal of Regulatory Economics 12 (3), 311–318.

Lee, S.-H. (1997b). "A note on regulating oligopolistic industries". Journal of Regulatory Economics 12 (1), 91–97.

Lee, S.-H., Hamilton, J. (1999). "Using market structure to regulate a vertically integrated monopolist". Journal of Regulatory Economics 15 (3), 223–248.

Lehman, D., Weisman, D. (2000). "The political economy of price cap regulation". Review of Industrial Organization 16 (4), 343–356.

Levine, P., Stern, J., Trillas, F. (2005). "Utility price regulation and time inconsistency: Comparisons with monetary policy". Oxford Economic Papers 57 (3), 447–478.

Levy, B., Spiller, P. (1994). "The institutional foundations of regulatory commitment: A comparative analysis of telecommunications". Journal of Law and Economics and Organization 10 (2), 201–246.

Lewis, T., Sappington, D. (1988a). "Regulating a monopolist with unknown demand". American Economic Review 78 (5), 986–998.

Lewis, T., Sappington, D. (1988b). "Regulating a monopolist with unknown demand and cost functions". RAND Journal of Economics 18 (3), 438–457.

Lewis, T., Sappington, D. (1989a). "Countervailing incentives in agency problems". Journal of Economic Theory 49 (2), 294–313.
Lewis, T., Sappington, D. (1989b). "Inflexible rules in incentive problems". American Economic Review 79 (1), 69–84.
Lewis, T., Sappington, D. (1989c). "An informational effect when regulated firms enter unregulated markets". Journal of Regulatory Economics 1 (1), 35–46.
Lewis, T., Sappington, D. (1989d). "Regulatory options and price cap regulation". RAND Journal of Economics 20 (3), 405–416.
Lewis, T., Sappington, D. (1990). "Sequential regulatory oversight". Journal of Regulatory Economics 2 (4), 327–348.
Lewis, T., Sappington, D. (1991a). "Incentives for monitoring quality". RAND Journal of Economics 22 (3), 370–384.
Lewis, T., Sappington, D. (1991b). "Oversight of long-term investment by short-lived regulators". International Economic Review 32 (3), 579–600.
Lewis, T., Sappington, D. (1992). "Incentives for conservation and quality improvement by public utilities". American Economic Review 82 (5), 1321–1340.
Lewis, T., Sappington, D. (1997). "Information management in incentive problems". Journal of Political Economy 105 (4), 796–821.
Lewis, T., Sappington, D. (1999). "Access pricing with unregulated downstream competition". Information Economics and Policy 11 (1), 73–100.
Lewis, T., Sappington, D. (2000). "Motivating wealth-constrained actors". American Economic Review 90 (4), 944–960.
Lewis, T., Yildirim, H. (2002). "Learning by doing and dynamic regulation". RAND Journal of Economics 33 (1), 22–36.
Liston, C. (1993). "Price-cap versus rate-of-return regulation". Journal of Regulatory Economics 5 (1), 25–48.
Lockwood, B. (1995). "Multi-firm regulation without lump-sum taxes". Journal of Public Economics 56 (1), 31–53.
Loeb, M., Magat, W. (1979). "A decentralized method for utility regulation". Journal of Law and Economics 22 (2), 399–404.
Luton, R., McAfee, P. (1986). "Sequential procurement auctions". Journal of Public Economics 31 (2), 181–195.
Lynch, J., Buzas, T., Berg, S. (1994). "Regulatory measurement and evaluation of telephone service quality". Management Science 40 (2), 169–194.
Lyon, T. (1991). "Regulation with 20–20 hindsight: 'Heads I win, tails you lose' ". RAND Journal of Economics 22 (4), 581–595.
Lyon, T. (1992). "Regulation with 20–20 hindsight: Least-cost rules and variable costs". Journal of Industrial Economics 40 (3), 277–289.
Lyon, T. (1996). "A model of sliding-scale regulation". Journal of Regulatory Economics 9 (3), 227–247.
Ma, C.-T.A. (1994). "Renegotiation and optimality in agency contracts". Review of Economic Studies 61 (1), 109–130.
Ma, C.-T.A., Moore, J., Turnbull, S. (1988). "Stopping agents from 'cheating' ". Journal of Economic Theory 46 (2), 335–372.
Maggi, G., Rodriguez-Clare, A. (1995). "On countervailing incentives". Journal of Economic Theory 66 (1), 238–263.
Mandy, D. (2000). "Killing the goose that laid the golden egg: Only the data know whether sabotage pays". Journal of Regulatory Economics 17 (2), 157–172.
Mandy, D. (2002). "TELRIC pricing with vintage capital". Journal of Regulatory Economics 22 (3), 215–249.
Mandy, D., Sharkey, W. (2003). "Dynamic pricing and investment from static proxy models". Review of Network Economics 2 (4), 403–439.
Manelli, A., Vincent, D. (1995). "Optimal procurement mechanisms". Econometrica 63 (3), 591–620.
Mankiw, N.G., Whinston, M. (1986). "Free entry and social inefficiency". RAND Journal of Economics 17 (1), 48–58.

Mansell, R., Church, J. (1995). Traditional and Incentive Regulation: Applications to Natural Gas Pipelines in Canada. The Van Horne Institute, Calgary.

Martimort, D. (1999). "Renegotiation design with multiple regulators". Journal of Economic Theory 88 (2), 261–293.

Matthews, S. (2001). "Renegotiating moral hazard contracts under limited liability and monotonicity". Journal of Economic Theory 97 (1), 1–29.

McAfee, R.P. (2002). "Coarse matching". Econometrica 70 (5), 2025–2034.

McAfee, R.P., McMillan, J. (1987a). "Auctions and bidding". Journal of Economic Literature 25 (2), 699–738.

McAfee, R.P., McMillan, J. (1987b). "Competition for agency contracts". RAND Journal of Economics 18 (2), 296–307.

McAfee, R.P., McMillan, J. (1996). "Analyzing the airwaves auction". Journal of Economic Perspectives 10 (1), 159–175.

McGuire, T., Riordan, M. (1995). "Incomplete Information and optimal market structure: Public purchases from private producers". Journal of Public Economics 56 (1), 125–141.

McMillan, J. (1994). "Selling spectrum rights". Journal of Economic Perspectives 8 (3), 145–162.

Meyer, M., Vickers, J. (1997). "Performance comparisons and dynamic incentives". Journal of Political Theory 105 (3), 547–581.

Milgrom, P. (1998). "Game theory and the spectrum auctions". European Economic Review 42 (3–5), 771–778.

Mirrlees, J. (1976). "Optimal tax theory: A synthesis". Journal of Public Economics 6, 327–358.

Mookherjee, D. (1984). "Optimal incentive schemes with many agents". Review of Economic Studies 51 (3), 433–446.

Mussa, M., Rosen, S. (1978). "Monopoly and product quality". Journal of Economic Theory 18 (2), 301–317.

Myerson, R. (1979). "Incentive compatibility and the bargaining problem". Econometrica 47 (1), 61–73.

Nalebuff, B., Stiglitz, J. (1983). "Prizes and incentives: Towards a general theory of compensation and competition". Bell Journal of Economics 14 (1), 21–43.

Neu, W. (1993). "Allocative inefficiency properties of price-cap regulation". Journal of Regulatory Economics 5 (2), 159–182.

Newbery, D. (1999). Privatization, Restructuring, and Regulation of Network Utilities. MIT Press, Cambridge, MA.

Otsuka, Y. (1997). "A welfare analysis of local franchises and other types of regulation: Evidence from the cable TV industry". Journal of Regulatory Economics 11 (2), 157–180.

Palfrey, T. (1983). "Bundling decisions by a multiproduct monopolist with incomplete information". Econometrica 51 (2), 463–484.

Palmer, K. (1991). "Diversification by regulated monopolies and incentives for cost-reducing R&D". American Economic Review 81 (2), 266–270.

Pint, E. (1992). "Price-cap versus rate-of-return regulation in a stochastic-cost model". RAND Journal of Economics 23 (4), 564–578.

Potters, J., Rockenbach, B., Sadrieh, A., van Damme, E. (2004). "Collusion under yardstick competition: An experimental study". International Journal of Industrial Organization 22 (7), 1017–1038.

Prager, R. (1989). "Franchise bidding for natural monopoly: The case of cable television in Massachusetts". Journal of Regulatory Economics 1 (2), 115–132.

Radner, R. (1981). "Monitoring cooperative agreements in a repeated principal–agent relationship". Econometrica 49 (5), 1127–1148.

Radner, R. (1985). "Repeated principal–agent relationships with discounting". Econometrica 53 (5), 1173–1198.

Ramakrishnan, R.T.S., Thakor, A.V. (1991). "Cooperation versus competition in agency". Journal of Law, Economics and Organization 7 (2), 248–283.

Ramsey, F. (1927). "A contribution to the theory of taxation". Economic Journal 37 (145), 47–61.

Reiffen, D. (1998). "A regulated firm's incentive to discriminate: A reevaluation and extension of Weisman's result". Journal of Regulatory Economics 14 (1), 79–86.

Reiffen, D., Schumann, L., Ward, M. (2000). "Discriminatory dealing with downstream competitors: Evidence from the cellular industry". Journal of Industrial Economics 48 (3), 253–286.
Rey, P., Salanié, B. (1996). "On the value of commitment with asymmetric information". Econometrica 64 (6), 1395–1414.
Rey, P., Tirole, J. (2007). "A primer on foreclosure". In: Armstrong, M., Porter, R. (Eds.), Handbook of Industrial Organization, vol. 3. North-Holland, Amsterdam (this volume).
Riordan, M. (1984). "On delegating price authority to a regulated firm". RAND Journal of Economics 15 (1), 108–115.
Riordan, M. (1996). "Contracting with qualified suppliers". International Economic Review 37 (1), 115–128.
Riordan, M. (2002). "Universal residential telephone service". In: Cave, M., Majumdar, S., Vogelsang, I. (Eds.), Handbook of Telecommunications Economics, vol. I. North-Holland, Amsterdam.
Riordan, M., Sappington, D. (1987a). "Awarding monopoly franchises". American Economic Review 77 (3), 375–387.
Riordan, M., Sappington, D. (1987b). "Information, incentives, and organizational mode". Quarterly Journal of Economics 102 (2), 243–263.
Riordan, M., Sappington, D. (1988). "Optimal contracts with public ex post information". Journal of Economic Theory 45 (1), 189–199.
Riordan, M., Sappington, D. (1989). "Second sourcing". RAND Journal of Economics 20 (1), 41–58.
Rob, R. (1986). "The design of procurement contracts". American Economic Review 76 (3), 378–389.
Rochet, J.-C., Stole, L. (2003). "The economics of multidimensional screening". In: Dewatripont, M., Hansen, L.P., Turnovsky, S. (Eds.), Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress. Cambridge Univ. Press, Cambridge, UK.
Rogerson, W. (1985). "Repeated moral hazard". Econometrica 53 (1), 69–76.
Rogerson, W. (2003). "Simple menus of contracts in cost-based procurement and regulation". American Economic Review 93 (3), 919–926.
Ronnen, U. (1991). "Minimum quality standards, fixed costs, and competition". RAND Journal of Economics 22 (4), 490–504.
Salant, D. (1995). "Behind the revolving door: A new view of public utility regulation". RAND Journal of Economics 26 (3), 362–377.
Salant, D. (2000). "Auctions and regulation: Reengineering of regulatory mechanisms". Journal of Regulatory Economics 17 (3), 195–204.
Salant, D., Woroch, G. (1992). "Trigger price regulation". RAND Journal of Economics 23 (1), 29–51.
Sand, J. (2004). "Regulation with non-price discrimination". International Journal of Industrial Organization 22 (8), 1289–1307.
Sappington, D. (1980). "Strategic firm behavior under a dynamic regulatory adjustment process". Bell Journal of Economics 11 (1), 360–372.
Sappington, D. (1982). "Optimal regulation of research and development under imperfect information". Bell Journal of Economics 13 (2), 354–368.
Sappington, D. (1983). "Limited liability contracts between principal and agent". Journal of Economic Theory 29 (1), 1–21.
Sappington, D. (1986). "Commitment to regulatory bureaucracy". Information Economics and Policy 2 (4), 243–258.
Sappington, D. (1994). "Designing incentive regulation". Review of Industrial Organization 9 (3), 245–272.
Sappington, D. (2002). "Price regulation and incentives". In: Cave, M., Majumdar, S., Vogelsang, I. (Eds.), Handbook of Telecommunications Economics, vol. I. North-Holland, Amsterdam.
Sappington, D. (2003). "Regulating horizontal diversification". International Journal of Industrial Organization 21 (3), 291–315.
Sappington, D. (2005a). "On the irrelevance of input prices for make-or-buy decisions". American Economic Review 95 (5), 1631–1638.
Sappington, D. (2005b). "Regulating service quality: A survey". Journal of Regulatory Economics 27 (2), 123–154.

Sappington, D., Sibley, D. (1988). “Regulating without cost information: The incremental surplus subsidy scheme”. International Economic Review 29 (2), 297–306.
Sappington, D., Sibley, D. (1990). “Regulating without cost information: Further observations”. International Economic Review 31 (4), 1027–1029.
Sappington, D., Sibley, D. (1992). “Strategic nonlinear pricing under price cap regulation”. RAND Journal of Economics 23 (1), 1–19.
Sappington, D., Sibley, D. (1993). “Regulatory incentive policies and abuse”. Journal of Regulatory Economics 5 (2), 131–141.
Sappington, D., Weisman, D. (1996a). Designing Incentive Regulation for the Telecommunications Industry. MIT Press, Cambridge, MA.
Sappington, D., Weisman, D. (1996b). “Revenue sharing in incentive regulation plans”. Information Economics and Policy 8 (3), 229–248.
Sappington, D., Weisman, D. (2005). “Self-sabotage”. Journal of Regulatory Economics 27 (2), 155–175.
Scarpa, C. (1998). “Minimum quality standards with more than two firms”. International Journal of Industrial Organization 16 (5), 665–676.
Schmalensee, R. (1989). “Good regulatory regimes”. RAND Journal of Economics 20 (3), 417–436.
Schmidt, K. (2000). “The political economy of mass privatization and the risk of expropriation”. European Economic Review 44 (2), 393–421.
Schwermer, S. (1994). “Regulating oligopolistic industries: A generalized incentive scheme”. Journal of Regulatory Economics 6 (1), 97–108.
Sen, A. (1996). “Termination clauses in long-term contracts”. Journal of Economics and Management Strategy 5 (4), 473–496.
Severinov, S. (2003). “Optimal organization: Centralization, decentralization or delegation?” Mimeo. Duke University.
Shleifer, A. (1985). “A theory of yardstick competition”. RAND Journal of Economics 16 (3), 319–327.
Sibley, D. (1989). “Asymmetric information, incentives, and price cap regulation”. RAND Journal of Economics 20 (3), 392–404.
Sibley, D., Weisman, D. (1998). “Raising rivals’ costs: The entry of an upstream monopolist into downstream markets”. Information Economics and Policy 10 (4), 451–470.
Sidak, G., Spulber, D. (1997). Deregulatory Takings and the Regulatory Contract. Cambridge Univ. Press, Cambridge, UK.
Sobel, J. (1999). “A reexamination of yardstick regulation”. Journal of Economics and Management Strategy 8 (1), 33–60.
Spence, M. (1975). “Monopoly, quality, and regulation”. Bell Journal of Economics 6 (2), 417–429.
Spiegel, Y. (1994). “The capital structure and investment of regulated firms under alternative regulatory regimes”. Journal of Regulatory Economics 6 (3), 297–320.
Spiegel, Y., Spulber, D. (1994). “The capital structure of regulated firms”. RAND Journal of Economics 25 (3), 424–440.
Spiegel, Y., Spulber, D. (1997). “Capital structure with countervailing incentives”. RAND Journal of Economics 28 (1), 1–24.
Spiller, P. (1990). “Politicians, interest groups, and regulators: A multiple-principals agency theory of regulation (or “let them be bribed”)”. Journal of Law and Economics 33 (1), 65–101.
Stefos, T. (1990). “Regulating without cost information: A comment”. International Economic Review 31 (4), 1021–1026.
Stole, L. (1994). “Information expropriation and moral hazard in optimal second-source auctions”. Journal of Public Economics 54 (3), 463–484.
Tangerås, T. (2002). “Collusion-proof yardstick competition”. Journal of Public Economics 83 (2), 231–254.
Taylor, C. (1995). “Digging for golden carrots: An analysis of research tournaments”. American Economic Review 85 (4), 872–890.
Tirole, J. (1986a). “Hierarchies and bureaucracies: On the role of collusion in organizations”. Journal of Law, Economics and Organization 2 (2), 181–214.

Tirole, J. (1986b). "Procurement and renegotiation". Journal of Political Economy 94 (2), 235–259.
Vickers, J. (1991). "Privatization and the risk of expropriation". Rivista Di Politica Economica 11, 115–146.
Vickers, J. (1995a). "Competition and regulation in vertically related markets". Review of Economic Studies 62 (1), 1–17.
Vickers, J. (1995b). "Concepts of competition". Oxford Economic Papers 47 (1), 1–23.
Vogelsang, I. (1988). "A little paradox in the design of regulatory mechanisms". International Economic Review 29 (3), 467–476.
Vogelsang, I. (1989). "Price cap regulation of telecommunications services: A long-run approach". In: Crew, M. (Ed.), Deregulation and Diversification of Utilities. Kluwer Academic Publishers, Boston, MA.
Vogelsang, I. (1990). "Optional two-part tariffs constrained by price caps". Economics Letters 33 (3), 287–292.
Vogelsang, I. (2002). "Incentive regulation and competition in public utility markets: A 20-year perspective". Journal of Regulatory Economics 22 (1), 5–28.
Vogelsang, I. (2003). "Price regulation of access to telecommunications networks". Journal of Economic Literature 41 (3), 830–862.
Vogelsang, I., Finsinger, J. (1979). "A regulatory adjustment process for optimal pricing by multiproduct monopoly firms". Bell Journal of Economics 10 (1), 151–171.
Weisman, D. (1993). "Superior regulatory regimes in theory and practice". Journal of Regulatory Economics 5 (4), 355–366.
Weisman, D. (1994). "Why more may be less under price-cap regulation". Journal of Regulatory Economics 6 (4), 339–362.
Weisman, D. (1995). "Regulation and the vertically integrated firm: The case of RBOC entry into InterLATA long distance". Journal of Regulatory Economics 8 (3), 249–266.
Weisman, D. (1998). "The incentive to discriminate by a vertically-integrated firm: A reply". Journal of Regulatory Economics 14 (1), 87–91.
Weisman, D. (2002). "Did the High Court reach an economic low in Verizon v. FCC?". Review of Network Economics 1 (2), 90–105.
Weisman, D. (2005). "Price regulation and quality". Information Economics and Policy 17 (2), 165–174.
Weisman, D., Williams, M. (2001). "The costs and benefits of long-distance entry: Regulation and non-price discrimination". Review of Economic Organization 18 (3), 275–282.
Williamson, O. (1975). Markets and Hierarchies: Analysis and Antitrust Implications. Free Press, New York.
Williamson, O. (1976). "Franchise bidding for natural monopolies – In general and with respect to CATV". Bell Journal of Economics 7 (1), 73–104.
Willig, R. (1979). "The theory of network access pricing". In: Trebing, H. (Ed.), Issues in Public Utility Regulation. Michigan State Univ. Press, East Lansing, MI.
Wilson, R. (1979). "Auctions of shares". Quarterly Journal of Economics 93 (4), 675–689.
Wolinsky, A. (1997). "Regulation of duopoly: Managed competition vs. regulated monopolies". Journal of Economics and Management Strategy 6 (4), 821–847.
Zupan, M. (1989a). "Cable franchise renewals: Do incumbent firms behave opportunistically?". RAND Journal of Economics 20 (4), 473–482.
Zupan, M. (1989b). "The efficacy of franchise bidding schemes in the case of CATV: Some systematic evidence". Journal of Law and Economics 32 (2), 401–456.