Recent Developments in the Theory of Regulation*

Mark Armstrong
Nuffield College, Oxford

David Sappington
Department of Economics, University of Florida

June 2003

Contents

*This is a draft of a chapter for the forthcoming *Handbook of Industrial Organization* (Vol. III). We are grateful to Carli Coetzee, Simon Cowan, Ernesto Dal Bó, Jos Jansen, Jean Tirole and Ingo Vogelsang for helpful comments.

1 Introduction

Several chapters in this volume analyze unfettered competition among industry producers. 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, 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 solely to pursue its own limited interests (e.g., profit maximization).

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 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 analysis of regulatory policies in a monopoly setting where the regulator's information 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.

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

[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 in this chapter. We offer our apologies in advance to the authors of any uncited contribution, appealing to limited knowledge and asymmetric 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.

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

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 operating technology and consumer demand. Consequently, it is important to analyze the optimal design of regulatory policy in settings that entail adverse selection (or “hidden information”) problems. This section reviews the relevant literature on this subject for the case where the regulated firm is a monopoly.[4]

Two distinct static settings of the regulation problem with adverse selection are considered in section 2.1. 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 2.1 concludes by presenting a unified framework for analyzing these various settings.

Section 2.2 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 for the possibility that the regulator is susceptible to capture by the industry, and to allow the firm’s private information to be multi-dimensional. Section 2.3 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.

Regulated firms also typically have better information about their actions than do regulators. Consequently, it is important to analyze the optimal design of regulatory policy in settings that entail moral hazard (or “hidden action”). Section 2.4 analyzes a particular regulatory moral hazard problem in which the firm’s cost structure is endogenous.

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

2.1 Regulation Under Adverse Selection

In this section we analyze simple versions of the central models of optimal regulation with private, but exogenous, information. 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 and more general framework.

We begin by defining the regulator's objective. We assume the regulator wishes to limit transfer payments from consumers/taxpayers to the firm because he values consumer surplus, S, more highly than the rent (or net profit) of the regulated firm. To capture this fact, we let $\alpha \in [0, 1]$ denote the value the regulator places on each dollar of the firm's rent, R, and assume that the regulator employs his policy instruments to maximize the expected value of $S + \alpha R$. The regulator's preference for consumer surplus over rent might reflect the deadweight loss that arises when consumers are taxed, for example. Alternatively, this preference might simply reflect a greater concern with the welfare of consumers than the welfare of shareholders. The extreme case where $\alpha = 1$ can be viewed as one in which the regulator values the welfare of consumers and shareholders equally, or in which consumer taxation entails no deadweight loss.[5]

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. In the full-information (or first-best) setting, the regulator will set the price(s) for the firm's product(s) equal to marginal production cost(s). Furthermore, when $\alpha < 1$, the firm's rent is socially costly, and so the regulator will ensure $R = 0$ by implementing the smallest transfer payment from consumers that ensures the firm is willing to operate. This ideal outcome for the regulator will be called the full-information outcome.

2.1.1 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 only 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 $p \geq 0$ 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 not fully informed about the level of (unobserved) fixed cost the firm must incur to realize any specified level of marginal cost.

In all three variants of this model, the regulator sets a unit price, p, for the regulated product. The regulator also specifies a transfer payment, T, from consumers to the regulated

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

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 from the regulator.

Unknown marginal cost[6]

For simplicity, suppose the firm produces with constant marginal cost that can take two values, $c \in \{c_L, c_H\}$. Let $\Delta^c = c_H - c_L > 0$ denote the cost differential between the high and the low marginal cost. The firm knows from the outset of its interaction with the regulator whether its marginal cost is 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 $F \geq 0$ 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 $i = L, H$.[7] When the firm with cost c_i chooses the (p_i, T_i) option, its rent will be

$$R_i = Q(p_i)(p_i - c_i) - F + T_i \ . \tag{1}$$

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 (i.e., the firm with the low marginal cost c_L) is to be induced to choose the (p_L, T_L) option, it must be the case that

$$R_L \geq R_H + \Delta^c Q(p_H) \ . \tag{2}$$

Therefore, the full-information outcome is not feasible, since inequality (2) cannot hold when both $R_H = 0$ and $R_L = 0$.[8]

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. (A sketch of the proofs of Propositions 1 to 4 is provided in section 2.1.3 below.)

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

$$p_L = c_L \ ; \ p_H = c_H + \frac{\phi}{1-\phi}(1-\alpha)\Delta^c \ ; \tag{3}$$

$$R_L = \Delta^c Q(p_H) \ ; \ R_H = 0 \ . \tag{4}$$

[6] 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 nonlinear costs, and where the firm's private information is the realization of a continuous random variable.

[7] 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).

[8] This conclusion assumes $Q(c_H) > 0$, which will be true as long as the marginal value of the initial level of output sufficiently exceeds even the highest marginal production cost. This will be assumed to be the case throughout the ensuing discussion, unless otherwise noted.

As expression (4) reveals, the regulator optimally provides the low-cost firm with the minimum possible rent require 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 (2) above.) Although the increase in p_H above c_H reduces the rent of the low-cost firm—which serves to increase welfare—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 (3) 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 (3) states that the regulator always 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 never has an incentive to choose the (p_L, T_L) option. As expression (4) 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 (3) 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 conclusion holds more generally as long as the regulator knows how consumers value the firm's output.[9] To see this, write $v(p)$ for consumer surplus when the price is p, and write $\pi(p)$ for the firm's profit function (a function that could be known only by the firm). Then the regulator could promise the firm a transfer of $T = v(p)$ when it sets the price p. Under this reward structure, the firm would choose 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 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, but this distribution is acceptable in the special case where the regulator cares only about total surplus. (However, in section 3.2.2 we will see how, in a dynamic context, surplus can sometimes be returned to consumers over time.) This conclusion—which is derived in an adverse selection setting—parallels the standard result in moral hazard principal-agent framework that the full-information outcome can be achieved 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 do not matter ($\alpha = 1$). The moral hazard problem is analyzed in section 2.4 below.

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

Countervailing incentives[10]

In the foregoing setting, the firm's natural 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, but 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.[11] 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 $\Delta^F \equiv F_L - F_H > 0$ denote the amount by which the firm's fixed cost of production increases as its marginal cost declines from to c_H to c_L. As before, let $\Delta^c \equiv c_H - c_L > 0$.

One might suspect that the regulator would suffer 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 cost 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* $p_L = c_L$;[12]
(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 potential variation in variable cost when marginal cost pricing is implemented. The usual incentive of the firm to exaggerate its marginal cost does 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 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 cost 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

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

[11] If fixed costs increased as marginal costs increased, the firm would have *added* 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 the Proposition 1 persist in this setting.

[12] 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 (3), 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 $(\Delta^c Q(\hat{p}_H), \Delta^c Q(c_H))$.

variation is less pronounced, then part (ii) of the result demonstrates that the qualitative distortions identified in Proposition 1 arise. 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 at a rate of compensation that is unprofitable 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 highest marginal cost is reported (i.e., $p_H = c_H$).

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).[13]

Unknown scope for cost reduction[14]

Now consider a setting where the regulator can observe the firm's marginal cost, but where this cost is chosen by the firm (rather than being chosen exogenously by nature). The regulator is uncertain about the unobserved effort (which we model as a fixed cost) required to achieve any given level of marginal cost. The regulator's limited information enables the firm to earn positive rent when the firm finds it easy to achieve low marginal cost. To limit these rents, the regulator limits the firm's reward for low realized production cost. Consequently, the firm generally does not reduce marginal cost to its efficient (full-information) level.

To characterize the optimal regulatory policy in this setting more precisely, suppose that 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 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 $1 - \phi$. In contrast with the previous models, the regulator can observe the firm's realized marginal cost c in the present setting. However, he 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 level of marginal cost (c). Therefore, for each $i = L, H$

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

[14] This is a simplified version of the model proposed in Laffont and Tirole (1986) and chapters 1 and 2 in Laffont and Tirole (1993b). See also Sappington (1982).

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

$$R_i = Q(p_i)(p_i - c_i) - F_i(c_i) + T_i \; . \tag{5}$$

As in inequality (2) above, the constraint that the low-cost firm does not claim to have high cost is

$$R_L \geq R_H + F_H(c_H) - F_L(c_H) \; . \tag{6}$$

Net consumer surplus in state i is $v(p_i) - T_i$. Using equality (5), this net surplus can be written as

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

Notice that the regulator's choice of prices $\{p_L, p_H\}$ does not affect the binding incentive constraint (6), given the choice of rents $\{R_L, R_H\}$. Consequently, prices do not affect rents. Therefore, prices will be set equal to the realized marginal costs (i.e., $p_i = c_i$) in order to maximize consumer surplus in (7). This conclusion reflects Laffont and Tirole's "incentive-pricing dichotomy": prices (generally) should be used solely to attain allocative efficiency, while rents should be used to motivate the firm to produce at low cost.[15]

If the regulator knew the firm's type, he would also require the efficient level of marginal cost, which is the cost that maximizes total surplus $\{v(c) - F_i(c)\}$. As usual, though, the full-information outcome is not feasible when the regulator does not share the firm's knowledge of its technology. To limit the low-cost firm's rent, the regulator inflates the high-cost 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:*

$$p_L = c_L \; ; \; p_H = c_H \; ; \tag{8}$$

$$Q(c_L) + F'_L(c_L) = 0 \; ; \tag{9}$$

$$Q(c_H) + F'_H(c_H) = -\frac{\phi}{1-\phi}(1-\alpha)(F'_H(c_H) - F'_L(c_H)) > 0, \tag{10}$$

$$R_L = F_H(c_H) - F_L(c_H) > 0 \; ; \; R_H = 0 \; . \tag{11}$$

Expression (9) indicates that the type-L firm will be induced to operate with the cost-minimizing technology. In contrast, expression (10) shows that the type-H firm will produce with inefficiently high marginal cost. This high marginal cost limits the rent that accrues

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

to the type-L firm, which, from inequality (6), decreases as c_H increases. As expression (10) 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.[16]

2.1.2 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.[17]

The firm's cost function, $C(\cdot)$, is common knowledge, but consumer demand can take two forms: the demand function is $Q_L(p)$ with probability ϕ or $Q_H(p)$ with probability $1-\phi$, where $Q_H(p) > Q_L(p)$ for all prices. 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.[18] As above, 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 in this setting. 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 in both demand realizations, which is the full-information outcome. Because marginal cost is constant with this technology, the full-information pricing policy (i.e., $p = c$) is common knowledge and does not depend upon the firm's private information.[19]

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 $C''(q) \geq 0$ then the full-information outcome is feasible (and optimal);

[16] 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).

[17] The following discussion is based on Lewis and Sappington (1988a). Riordan (1984) analyzes a model where the firm's marginal cost is constant up to an endogenous capacity level. In Riordan's model, the firm learns demand only after choosing its capacity level, and is willing to operate whenever it anticipates nonnegative expected profit.

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

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

(ii) If $C''(q) < 0$ then the regulator often[20] *sets a single price and transfer 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 corresponding increase in the transfer payment it receives.[21] In sum, part (i) of Proposition 4 states that the full-information outcome is feasible in this setting.[22]

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.[23] 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.[24]

[20] The precise meaning of "often" is made clear in section 2.1.3. To illustrate, pooling is optimal when $Q'_L(p) = Q'_H(p)$ for all p, so that the two demand functions differ by an additive constant.

[21] 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. Lewis and Sappington (1992) show that this result continues to hold when the regulated firm chooses the level of observable and contractible quality that it supplies.

[22] 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.

[23] This will be the case when the marginal cost curve is "flatter" than the inverse demand curve, and so the problem is concave (and there exists a unique welfare-maximizing price that equals marginal cost in each state).

[24] A similar feature will emerge in section 2.3, which discusses a dynamic regulatory setting in which the

Notice that in the present setting where the regulator seeks to maximize a weighted sum of consumer surplus and profit, the relevant full-information benchmark is pricing at marginal cost. An alternative setting is where the regulator seeks to maximize total (unweighted) surplus, but where transfer payments from taxpayers to the firm are socially costly.[25] When a transfer payment to the firm imposes a deadweight loss on society, Ramsey prices, rather than marginal-cost prices, become the relevant full-information benchmark. Of course, implementation of Ramsey prices requires knowledge of consumer demand. Consequently, in the alternative setting 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 setting where transfer payments to the firm are socially costly. In contrast, the qualitative conclusions drawn in Propositions 1-3 persist when transfer payments are socially costly.

2.1.3 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.[26] 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 $\pi_i(p)$. Again, p_i is the firm's equilibrium unit price and T_i is the corresponding transfer payment from the regulator to the firm in state i. The firm's equilibrium rent in state i is therefore $R_i = \pi_i(p_i) + T_i$.

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

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 it 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.

[25] See Laffont and Tirole (1986) and Laffont and Tirole (1993b).

[26] This material is taken from Armstrong and Sappington (2003). Guesnerie and Laffont (1984) and Caillaud, Guesnerie, Rey, and Tirole (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.

with p. Formally,

$$\Delta^\pi(p) \equiv \pi_H(p) - \pi_L(p) \text{ and } \frac{d}{dp}\Delta^\pi > 0 \,. \tag{12}$$

The "increasing difference" property in expression (12) reflects the standard single crossing property.[27] Its role, as will be shown below, is to guarantee that the equilibrium price in state H is necessarily higher than in state L.

The regulator seeks to maximize the expected value of a weighted average of consumer surplus and rent. Consumer surplus in state i given price p is the surplus obtained from consuming the product at price p, which is denoted $v_i(p)$, minus the transfer, T_i, to the firm. Written in terms of rents $R_i = \pi_i(p_i) + T_i$, this weighted average in state i when price p_i is charged is

$$S_i + \alpha R_i = v_i(p_i) - T_i + \alpha(\pi_i(p_i) + T_i) = w_i(p_i) - (1-\alpha)R_i \,. \tag{13}$$

Here, $w_i(p) \equiv v_i(p) + \pi_i(p)$ denotes total unweighted surplus (the sum of consumer surplus and profit) 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

$$W = \phi\left\{w_L(p_L) - (1-\alpha)R_L\right\} + (1-\phi)\left\{w_H(p_H) - (1-\alpha)R_H\right\} \,. \tag{14}$$

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

$$R_i \geq 0 \;\text{ for } i = L, H \,. \tag{15}$$

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 that $R_i = 0$. This is the full-information benchmark. If the regulator does not know the state of the world, 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 (2) and (6) above, the regulator must ensure that the following incentive compatibility constraints are satisfied:

$$R_L \geq R_H - \Delta^\pi(p_H) \,, \tag{16}$$

$$R_H \geq R_L + \Delta^\pi(p_L) \,. \tag{17}$$

Adding inequalities (16) and (17) implies

$$\Delta^\pi(p_H) \geq \Delta^\pi(p_L) \,. \tag{18}$$

The increasing difference assumption in expression (12) together with inequality (18) imply that the equilibrium price must be higher in state H than in state L in any incentive-compatible regulatory policy, i.e.,

$$p_H \geq p_L \,. \tag{19}$$

[27] 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.

A key conclusion that aids in understanding the solution to the regulator's problem is the following.

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.,* $p_j = p_j^*$.[28]

To understand this result, suppose that, say, the incentive compatibility constraint for the type H firm, inequality (17), does not bind at the optimum. Then, keeping 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 (17). If a small change in p_L does not increase welfare $w_L(p_L)$ in (14), 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.[29] The incentive constraints (16) and (17) imply that this full-information outcome is attainable when the regulator does not observe the state if and only if

$$\Delta^\pi(p_H^*) \geq 0 \geq \Delta^\pi(p_L^*) \ . \tag{20}$$

The pair of inequalities in (20) 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 constant 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:

$$C(Q_i(p_j^*)) \geq C(q_j^*) + C'(q_j^*)(Q_i(p_j^*) - q_j^*). \tag{21}$$

To show that inequality (20) is satisfied when prices are equal to marginal costs, notice that

$$\begin{aligned} \pi_i(p_j^*) &= p_j^* Q_i(p_j^*) - C(Q_i(p_j^*)) \\ &\leq p_j^* Q_i(p_j^*) - \left\{C(q_j^*) + C'(q_j^*)(Q_i(p_j^*) - q_j^*)\right\} \\ &= p_j^* q_j^* - C(q_j^*) \\ &= \pi_j(p_j^*). \end{aligned} \tag{22}$$

[28] This assumes that the surplus functions $w_i(p_i)$ are single-peaked.

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

The inequality in expression (22) follows from inequality (21). The second equality in expression (22) holds because $p_j^* = C'(q_j^*)$. Consequently, condition (20) 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 (20) are satisfied. Since $\pi_i(p) = (p - c_i)Q(p) - F_i$, it follows that

$$\Delta^\pi(p) = \Delta^F - \Delta^c Q(p) \ . \tag{23}$$

Therefore, since full-information prices are $p_i^* \equiv c_i$, expression (23) implies that the inequalities (20) will be satisfied if and only if $\Delta^c Q(c_L) \geq \Delta^F \geq \Delta^c Q(c_H)$, as indicated in Proposition 2.

Price distortions with separation

When profit is higher in state L than in state H for any specified price (as, for example, when the firm is privately informed only about its constant marginal cost of production), then $\Delta^\pi(p) < 0$ for all p. In this case, only the type-H firm's participation constraint in (15) will be relevant, and this firm will optimally be afforded no rent. (Since rents have social cost in the formulation in expression (14), and the incentive compatibility constraints (16)–(17) depend only on the *difference* between the rents, it is clear that at least one participation constraint must bind at the optimum.) In this case, the incentive compatibility constraints (16)–(17) become

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

Again, since rent is socially costly, it is only the *lower* bound on R_L that is relevant, i.e., only the incentive compatibility constraint (16) is relevant.

Therefore, expression (14) reduces to

$$W = \phi\{w_L(p_L) + (1-\alpha)\Delta^\pi(p_H)\} + (1-\phi)w_H(p_H) \ , \tag{24}$$

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

Maximizing expression (24) with respect to p_L and p_H implies:

$$p_L = p_L^* \ \text{ and } p_H \text{ maximizes } w_H(p) + \frac{\phi}{1-\phi}(1-\alpha)\Delta^\pi(p) \ . \tag{25}$$

Since $\Delta^\pi(p)$ increases with p, expression (25) implies that $p_H > p_H^*$. When full-information prices are ordered as in inequality (19) it follows that $p_H > p_H^* \geq p_L^* = p_L$, and therefore the monotonicity condition (19) is indeed satisfied. Therefore, (25) 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).

This analysis is presented in Figure 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 $R_L \geq -\Delta^\pi(p_H)$, and the regulator must limit himself to a contract that lies within this set. Expression (14) shows that iso-welfare contours in (p_H, R_L) space take the form $R_L = \left[\frac{1}{1-\alpha}\right]\left[\frac{1-\phi}{\phi}\right] w_H(p_H) + k_o$, where k_o is a constant. Each of these contours is maximized at $p_H = p_H^*$ as depicted on the diagram. (Lower contours yield higher welfare.) Therefore, the optimum is where an iso-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 iso-welfare contours, and so brings the optimal choice of price p_H closer to the full-information level.

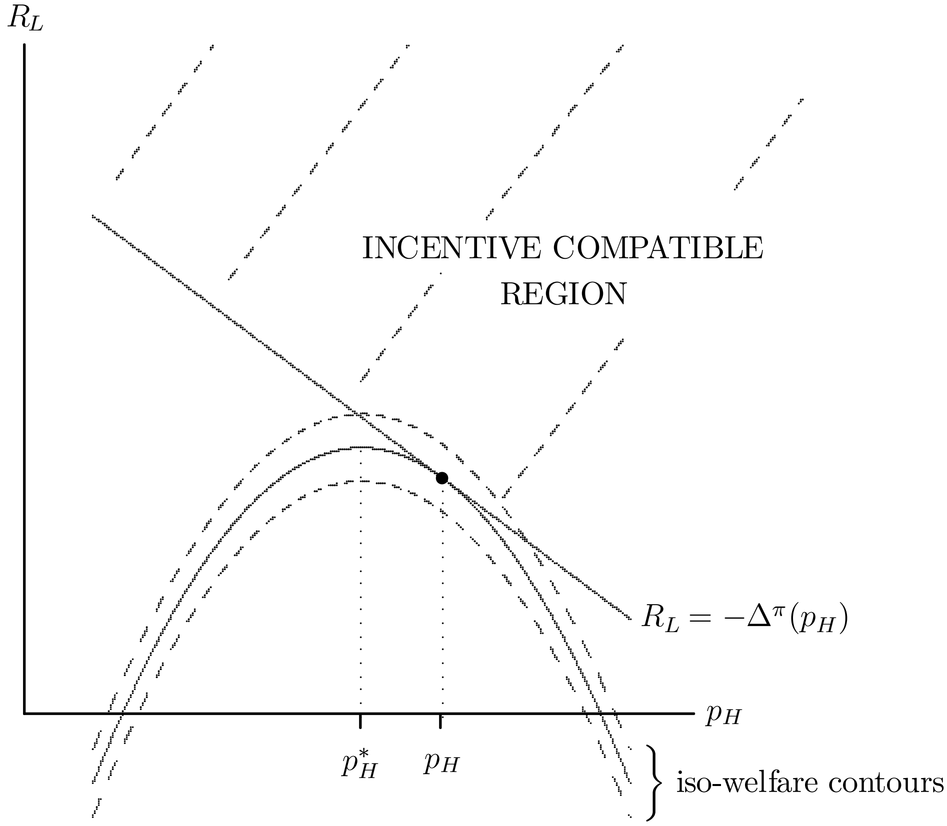

Figure 1: Price Distortions with Separation

These qualitative features characterize the optimal regulatory policy in many settings, including the setting of Proposition 1 where the firm is privately informed about (only) 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 (25) implies that the optimal price for the high-cost firm is as given in expression (3) of Proposition 1.

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 ($p_i = c_i$), 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 (25) yields expression (10) in Proposition 3.

In concluding this discussion of optimal separating prices, notice that welfare in expression (24) can be written as:

$$W = \phi\left\{v(p_L) + Q(p_L)(p_L - c_L)\right\} + (1-\phi)\left\{v(p_H) + Q(p_H)(p_H - \hat{c}_H)\right\} \ , \tag{26}$$

where

$$\hat{c}_H = c_H + \frac{\phi}{1-\phi}(1-\alpha)\Delta^c \ . \tag{27}$$

(Notice that $\hat{c}_H$ is simply p_H in expression (3).[30]) Expression (26) 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 realized 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 (27) in a setting with no information asymmetry. Under this interpretation, the prices in expression (3) are simply marginal-cost prices, where the "costs" are adjusted away from the underlying costs to take account of socially undesirable rents.

Pooling

It remains to illustrate why the regulator might sometimes implement the same price in both states of the world. As suggested in the discussion after Proposition 4, pooling (i.e., $p_H = p_L$) may well be optimal if $p_L^* > p_H^*$, so that prices in the full-information outcome do not satisfy the necessary condition for incentive compatibility, which is inequality (19).

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. To ensure the regulator's problem is concave, the marginal cost curve is assumed to be "flatter" than the relevant inverse demand curve, so that $|C''(Q_i(p))| < 1/|Q_i'(p)|$ for $i = L, H$.[31] Because marginal cost declines with output in this setting, the full-information prices satisfy $p_L^* > p_H^*$.

It is useful to restrict attention to cases where the single-crossing condition is satisfied. In this case with unknown demand, condition (12) is satisfied when

$$pQ_H'(p) + Q_H(p) - C'(Q_H(p))Q_H'(p) > pQ_L'(p) + Q_L(p) - C'(Q_L(p))Q_L'(p) \ . \tag{28}$$

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 (28) is satisfied (provided the concavity condition is also satisfied).

To verify 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 (16) or (17) binds. Therefore, from Lemma 1, the

[30] The adjusted cost $\hat{c}_H$ would be refered to as the "virtual cost" within the mechanism design literature.

[31] This condition ensures there exists a unique price that equals realized marginal cost in each state, and that this price maximizes welfare.

full-information price p_i^* is implemented in one state.[32] Suppose state L is the relevant state, so that $p_L = p_L^*$ and (16) binds:

$$R_L = R_H - \Delta^\pi(p_H) \ . \tag{29}$$

Also, an analysis analogous to that which underlies expressions (21) and (22) reveals that $\pi_H(p_L^*) \geq \pi_L(p_L^*)$ when $C(\cdot)$ is concave, i.e., $\Delta^\pi(p_L^*) \geq 0$. Since inequality (19) requires $p_H > p_L^*$ and since Δ^π is increasing, it follows that $\Delta^\pi(p_H) > 0$. Since at least one participation constraint (15) will bind, expression (29) then implies

$$R_L = 0 \ ; \ \ R_H = \Delta^\pi(p_H) \ . \tag{30}$$

Therefore, expected welfare in expression (14) simplifies to

$$W = \phi w_L(p_L^*) + (1-\phi)\left\{w_H(p_H) - (1-\alpha)\Delta^\pi(p_H)\right\} \ . \tag{31}$$

Since $p_H > p_H^*$ it follows that the term in $\{\cdot\}$ brackets in equation (31) is decreasing in p_H provided $w_H(\cdot)$ is single-peaked in price. Since a small reduction in p_H does not violate any participation or incentive compatibility constraint, and will increase the value of the regulator's objective function, the candidate prices $\{p_L^*, p_H\}$ cannot be optimal. A similar argument holds if inequality (17) is the binding incentive constraint. Therefore, by contradiction, $p_L = p_H$ in the solution to the regulator's problem.[33]

Notice that, unlike the pricing distortions discussed above (e.g., in expression (25)), 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 concerns.

2.2 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 2.1.1, where the demand function facing the firm is common knowledge.

---

[32] If the single-crossing condition holds, both incentive constraints can only bind when $p_L = p_H$. On the other hand, if neither constraint binds, then it follows that $p_L = p_L^*$ and $p_H = p_H^*$, which cannot be incentive compatible when $p_L^* > p_H^*$.

[33] For further discussion of when pooling will arise in models of this sort, see section 2.10.2 of Laffont and Martimort (2002).

2.2.1 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 in this setting 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 could be interpreted as being the output of an "audit" of the firm's cost. 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 underlying cost, assume that $\phi_L > \phi_H$.[34]

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 unit price the firm selects and on the realization of the public signal. Specifically, let T_{ij} be the regulator's transfer to the firm when it claims its cost is c_i and when the realized signal turns out to 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 of F to the firm when its marginal cost is indeed c_H. Formally:

$$\phi_H T_{HL} + (1 - \phi_H) T_{HH} = F \ . \tag{32}$$

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 (c_H). Formally:

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

The transfers T_{HL} and T_{HH} can always be set to satisfy equality (32) and inequality (33) except in the case where the signal is entirely uninformative ($\phi_L = \phi_H$). 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.[35]

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.[36] To limit these rents, the regulator will implement the qualitative pricing distortions identified in Proposition 1. 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.[37]

[34] Another way to model this "audit" would be to suppose that with some exogenous probability the regulator observes the true cost (and otherwise observes "nothing"). This alternative specification yields the same insights. A form of this alternative specification is explored in the next section on regulatory capture.

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

[36] See Baron and Besanko (1984b) for this analysis.

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

An interesting extension of this analysis is when the regulator has to incur a cost to receive the audit, and therefore has to decide *when* to purchase the signal.[38] 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.[39] The frequency of this audit is determined by balancing the costs of auditing with the benefits of improved information.

2.2.2 Partially Informed Regulator: Regulatory Capture

In this section we relax the assumption that the regulator automatically acts in the interests of society.[40] Indeed, for simplicity we take the other extreme and suppose that 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 reveals this damaging information. This incentive scheme is designed by a "political principal", who might be viewed as the (benevolent) government, for example.[41]

To be specific, suppose the firm can have two levels of marginal cost, c_L and c_H, where the probability of a low cost realization is ϕ. Also, suppose that conditional on the firm's cost realization being low, the regulator has an exogenous chance ζ of being informed that the cost is indeed low. Conditional on a high cost realization, the regulator has no chance of being informed.[42] 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

[38] See Baron and Besanko (1984b) and section 3.6 of Laffont and Martimort (2002) for this analysis.

[39] 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.

[40] This discussion is based on Laffont and Tirole (1991b) and Chapter 11 of Laffont and Tirole (1993b). 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 not on collusion but on giving the regulator good incentives 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.

[41] An alternative formulation is that the regulator commissions an auditor to gather information about the firm. The firm might then try to bribe the auditor not to reveal detrimental information to the regulator.

[42] Chapter 11 of Laffont and Tirole (1993b) models the information structure more symmetrically in that the regulator has a chance ζ of being informed about the true cost, regardless of whether it is high or low. However, when the regulator learns that the cost is high, the firm has no interest in persuading him to conceal this information. Since the possibility that the regulator is informed that costs are high plays no significant role in this analysis of capture, but does make the notation more cumbersome, we assume the regulator can obtain information only about a low cost realization.

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 probability that the cost is low, conditional on the regulator being informed, is 1.)

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, there is no way the political principal can confirm this is in fact the case. The political principal is unable to detect 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.[43] Suppose also that the regulator is paid an extra amount s when he admits to being informed. Assume for now that the political principal decides to pay the regulator enough to make it in his interest to reveal his information when he is informed, i.e., that the principal restricts attention to "collusion-proof" mechanisms. 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.

Suppose that it costs the firm \$$(1+\theta)$ to increase the income of the regulator by \$1. The extra marginal cost θ of 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 when other services are regulated more leniently. For simplicity, we model these indirect ways of influencing the regulator's decision as an extra marginal cost θ that the firm incurs in delivering income to the regulator. For expositional ease, we will speak of the firm as "bribing" the regulator, even when explicit bribery is not undertaken.[44]

It is clear from Proposition 1 that at the optimum the low-cost firm will 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 that $R_H = 0$, expression (2) implies that the rent of the low-cost firm (again, conditional on the regulator being uninformed) is $\Delta^c Q(p_H)$.

Let s denote the payment from the political principal to the regulator when the latter reports to have learned that the firm has low cost. The low-cost firm will find it too costly

[43] 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 addition, by normalizing the regulator's required income to zero, we introduce the implicit assumption that the regulator is somehow indispensable for regulation, and the political principal cannot do without his services and cannot avoid paying him his reservation wage.

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

to bribe the informed regulator to conceal his information if

$$(1+\theta)s \geq \Delta^c Q(p_H) . \tag{34}$$

Expression (34) is the incentive compatibility constraint which ensures that the corruptible regulator is truthful when he announces he is ignorant about the firm's cost.

Suppose that the regulator's income receives weight $\alpha_R \leq 1$ in the political principal's welfare function, while the rent of the firm has weight α. Then, analogous to expression (14), total expected welfare under this "collusion-proof" regulatory policy is

$$W = \psi \left[w_L(c_L) - (1-\alpha_R)s \right] +$$

$$(1-\psi) \left[\phi^U \{ w_L(c_L) - (1-\alpha)R_L \} + (1-\phi^U) w_H(p_H) \right] .$$

Since payments from the political principal to the regulator are socially costly, the regulator's incentive compatibility constraint in inequality (34) will bind at the optimum. Consequently, total expected welfare is

$$W = \psi \left[w_L(c_L) - \frac{1-\alpha_R}{1+\theta} \Delta^c Q(p_H) \right] +$$

$$(1-\psi) \left[\phi^U \{ w_L(c_L) - (1-\alpha)\Delta^c Q(p_H) \} + (1-\phi^U) w_H(p_H) \right] . \tag{35}$$

Before finding the price p_H that maximizes expected welfare, we can check when the political principal will design the reward structure to ensure that the regulator is not captured, i.e., when it is optimal to satisfy inequality (34). If the principal does not choose s to satisfy (34), 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 that the principal can do is to follow the Baron-Myerson regulatory policy described in Proposition 1, where the policy designer has no extra private information. From expression (24), expected welfare in this case is

$$W = \phi \{ w_L(c_L) - (1-\alpha)\Delta^c Q(p_H) \} + (1-\phi) w_H(p_H) . \tag{36}$$

Using the identity $\psi + (1-\psi)\phi^U = \phi$, a comparison of welfare in (35) and (36) shows 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 be susceptible to bribes from the firm—whenever $(1+\theta)(1-\alpha) > 1-\alpha_R$. In particular, whenever the regulator's rent receives as least as much weight in social welfare as the firm's rent, it is optimal to make use of the regulator's information. Assume for the remainder of this section that this inequality holds.

Maximizing expression (35) with respect to p_H yields:

$$p_H = \underbrace{c_H + \frac{\phi^U}{1-\phi^U}(1-\alpha)\Delta^c}_{\text{Baron-Myerson price}} + \underbrace{\frac{\psi}{(1-\psi)(1+\theta)(1-\phi^U)}(1-\alpha_R)\Delta^c}_{\text{extra distortion to reduce firm's stake in collusion}} . \tag{37}$$

From expression (3) 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 (37). The second term in (37) 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) when it is very costly for the firm to bribe the regulator (i.e., when $\theta = \infty$).

The reason why the price for the high-cost firm is distorted further above cost when capture is possible is that, from expression (34), a more distorted 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, therefore, 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 brief discussion has considered what one might term the "optimal response" to the danger of capture and collusion.[45] We return to the general topic in section 3.4.2, which focuses more on pragmatic responses to capture, such as restricting the regulator's discretion over his policy.

2.2.3 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 about consumer demand.

To analyze this situation formally, first consider the following simple multiproduct extension of the Baron and Myerson (1982) model described in section 2.1.1.[46] Suppose that 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 for product k is $c_k \in \{c_L, c_H\}$. 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 has cost c_i for product 1 (sold in market 1) and cost c_j for product 2 (sold in market 2). Suppose that, as in section 2.3.1, the unconditional probability that the firm has a low cost realization in market 1 is ϕ. Let ϕ_i be the probability that the firm has a low cost realization in market 2, given that its cost

[45] 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 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".

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

is c_i in market 1. The cost realizations are positively correlated across markets if $\phi_L > \phi_H$, negatively correlated if $\phi_L < \phi_H$, and statistically independent if $\phi_L = \phi_H$. In order to keep the analysis simple, suppose that the unconditional probability of a low cost realization in market 2 is also ϕ. In this case, states LH and HL are equally likely, so:

$$\phi(1-\phi_L) = (1-\phi)\phi_H \ .$$

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_{ij}^1 in market 1, p_{ij}^2 in market 2, and in return receive the transfer T_{ij}. Consequently, as with expression (1) in the single-product case, the equilibrium rent of the type ij firm is

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

The participation constraints in the regulator's problem take the form $R_{ij} \geq 0$, of which only $R_{HH} \geq 0$ is relevant. (If the firm is one of the other three types, it can claim to have high cost in both markets, 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 "downward" constraints, which ensure that low-cost types will not claim to have high costs.[47] 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 in one market and low in 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 in both markets when the firm has low cost in both markets; p_{HH} is the price in both markets when the firm has a high cost in both markets; p_L^A is the price in the low-cost market when the firm has asymmetric costs, while p_H^A is the price in the high-cost market when the firm has asymmetric costs.

Much like the single-product case in expression (14), equilibrium expected welfare in this multi-dimensional setting is:

$$W = 2\phi(1-\phi_L)\left\{w_L(p_L^A) + w_H(p_H^A) - (1-\alpha)R_A\right\} +$$

$$\phi\phi_L\{2w_L(p_{LL}) - (1-\alpha)R_{LL}\} + (1-\phi)(1-\phi_H)\{2w_H(p_{HH}) - (1-\alpha)R_{HH}\} \ . \tag{38}$$

(Here, $w_i(p) = v(p) + Q(p)(p - c_i)$, where $v(\cdot)$ again denotes consumer surplus.) And much like the single-product case in expression (2), the incentive compatibility constraint ensuring that the type-A firm does not claim to be the type-HH firm is:

$$\begin{aligned} R_A &\geq Q(p_{HH})(p_{HH} - c_H) + Q(p_{HH})(p_{HH} - c_L) - F + T_{HH} \\ &= R_{HH} + \Delta^c Q(p_{HH}) \ , \end{aligned} \tag{39}$$

[47]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.

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

$$R_{LL} \geq R_A + \Delta^c Q(p_H^A) \; . \tag{40}$$

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

$$R_{LL} \geq R_{HH} + 2\Delta^c Q(p_{HH}) \; . \tag{41}$$

The participation constraint for the type-HH firm will bind, so $R_{HH} = 0$. The type-A firm's incentive compatibility constraint (39) will also bind, so $R_A = \Delta^c Q(p_{HH})$. Substituting these values into (40) and (41) implies that the rent of the type-LL firm is

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

Substituting these rents into expected welfare (38) implies that welfare is

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

$$\phi\phi_L\{2w_L(p_{LL}) - (1-\alpha)2\Delta^c Q(p_{HH})\} + (1-\phi)(1-\phi_H)2w_H(p_{HH}) \tag{42}$$

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

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

$$\phi\phi_L\{2w_L(p_{LL}) - (1-\alpha)\Delta^c\left[Q(p_{HH}) + Q(p_H^A)\right]\} + (1-\phi)(1-\phi_H)2w_H(p_{HH}) \tag{43}$$

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

The policy that maximizes welfare consists of the prices $\{p_{LL}, p_{HH}, p_L^A, p_H^A\}$ that maximize the expression in (42)–(43). 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.[48] Second, the strict inequality $p_H^A > p_{HH}$ cannot be optimal. To see why notice that when this inequality holds, expression (42) is the relevant expression for welfare. In expression (39), the price p_H^A does not affect any rents. Consequently, $p_H^A = c_H$ is optimal. But the value of p_{HH} that maximizes expression (42) is strictly above cost c_H. Therefore, the inequality $p_H^A \geq p_{HH}$ must *bind* if (42) is maximized subject to the constraint $p_H^A \geq p_{HH}$. In sum, attention can be restricted to the case where $p_H^A \leq p_{HH}$, and so (43) 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 $p_H^A \leq p_{HH}$ is ignored, the prices that maximize (43) are:

$$p_H^A \text{ maximizes } 2w_H(\cdot) - \frac{\phi_L}{1-\phi_L}(1-\alpha)\Delta^c Q(\cdot) \; ; \text{ and} \tag{44}$$

[48] 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.

$$p_{HH} \text{ maximizes } 2w_H(\cdot) - \frac{1-(1-\phi)(1-\phi_H)}{(1-\phi)(1-\phi_H)}(1-\alpha)\Delta^c Q(\cdot) \ . \tag{45}$$

Clearly, the price p_{HH} that solves (45) is higher that the price p_H^A that solves (44) whenever

$$\phi_L \leq 1-(1-\phi)(1-\phi_H) \ ,$$

which, after some rearrangement, is equivalent to the condition $\phi_L \leq 2\phi_H$. This inequality states that there is "not too much" positive correlation between cost realizations in the two markets. When this condition is satisfied, expressions (44) and (45) give the two high-cost prices.

When there is strong positive correlation, so $\phi_L \geq 2\phi_H$, the constraint $p_H^A \leq p_{HH}$ binds. Letting $p_H = p_H^A = p_{HH}$ denote this common price, (43) simplifies to

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

which is just (twice) the standard single-product Baron–Myerson formula. (See expression (24) for instance.) Therefore, with strong positive correlation, the solution is simply (two copies of) the Baron-Myerson formula (3). This discussion constitutes the proof of Proposition 5.

Proposition 5 *The optimal policy in the symmetric multi-dimensional setting has the following features:*
(i) There are no pricing distortions in markets where costs are low, i.e., $p_{LL} = p_L^A = c_L$.
(ii) When there is strong positive correlation between costs (so $\phi_L \geq 2\phi_H$*), regulatory policy in each market is independent of the firm's report for the other market. The policy in each market is identical to the policy described in Proposition 1.*
(iii) When cost correlation is weak (so $\phi_L \leq 2\phi_H$*), interdependencies are introduced across markets. In particular:*

$$p_H^A = c_H + \frac{\phi_L}{2(1-\phi_L)}(1-\alpha)\Delta^c \ , \ \textit{and}$$

$$p_{HH} = c_H + \frac{1-(1-\phi)(1-\phi_H)}{2(1-\phi)(1-\phi_H)}(1-\alpha)\Delta^c \geq p_H^A \ .$$

Part (i) of Proposition 5 provides the standard conclusion that price is set equal to cost when the firm has low cost in a market. 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 in each market and depends only on the cost realized in that market. 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.

Property (iii) of Proposition 5 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 high cost in one market, its price is set closer to cost in that market when its cost is low in the other market than when its cost is high in the other market. 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.[49] In contrast, simultaneous high cost realizations in both markets are 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 in both markets in the relatively likely event that the firm has high cost in one market and low cost in 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 consumer demand for its product.[50] 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.[51]

2.3 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 2.1.1. For simplicity, suppose that the firm's cost $c \in \{c_L, c_H\}$ is perfectly correlated across the two periods.[52]. Let $\phi \in (0,1)$ be the probability that the firm has low marginal cost, c_L, for the two periods. The regulator and the firm have the common 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 will receive non-negative rent

[49] Notice from property (iii) of Proposition 5 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.

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

[51] 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 (as it is here) or continuous. One reason for the difference is that in a continuous framework it is generally optimal to *shut down* some firms in order to extract further rent from the remainder, a feature that tends to complicate the analysis. See section 2 of Armstrong (1999) for further analysis of this issue. See Rochet and Stole (2003) for a survey of multidimensional screening.

[52] See Baron and Besanko (1984a) and section 8.1.3 in Laffont and Martimort (2002) 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.

from doing so, just as it will only produce in the first period if it anticipates non-negative expected discounted rent from doing so.

In the ensuing sections we analyze formally 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.

2.3.1 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.

Proposition 6 *In the two-period setting with perfect intertemporal regulatory commitment, the optimal regulatory policy has the following features:*
(i) Prices in the two periods are

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

(ii) Total rents are

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

Thus, the regulator implements the same (Baron-Myerson) pricing policy in each period.

Once the regulator has observed the choice made by the firm in the first period, he would wish at that point—if he were free to do so—to change second-period policy in one of two ways. Recall from Proposition 1 that there are two things that, when compared to the full-information outcome, are undesirable about the optimal regulatory policy with asymmetric information. First, the high-cost firm charges a price that is distorted above its marginal cost (but this firm has no rents). Second, the low-cost firm obtains a socially costly rent (but sets price equal to marginal cost). By the second period, the regulator has full information about the firm's cost. Therefore, if the firm reveals that it has high cost, the regulator would, at that point in time, prefer to reduce the firm's price to the level of its cost. Here, the temptation is not so much to eliminate rents (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 reveals that 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 would like to expropriate the firm. Such a change in regulatory policy would not be mutually improving. These two temptations are the subject of the two kinds of commitment problems discussed next.

When there is full commitment power, however, the regulator does not succumb to these temptations, and he makes a commitment not to use against the firm in the second period any cost information he infers from its first-period actions. He does this in order best to limit the rent that accrues to the firm with low cost.

2.3.2 Long-term Contracts: The Danger of Renegotiation

Now consider the case where the regulator has "moderate" commitment powers.[53] 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 (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, in the light of information revealed to him, to be Pareto inefficient. 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 rents to the firm, but not to *how* those rents are generated (i.e., to the particular prices and transfers that generate the rent). The firm does not care how its rents are generated, but the composition of rent does affect the firm's incentives to reveal its cost truthfully.

In this framework, 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 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.[54] If he proposes a change, the firm then decides whether to accept the change. 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, the total discounted

[53] This discussion in this section is based on Laffont and Tirole (1990a) and chapter 10 of Laffont and Tirole (1993b). For an alternative model of "moderate" commitment power, see Baron and Besanko (1987a).

[54] 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 pages 443–447 of Laffont and Tirole (1993b) for further discussion of this issue.

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. Since 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 T 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

$$\begin{aligned} R_L &\geq Q(p_H)(p_H - c_L) - F + T_H + \delta \left\{ Q(c_H)(c_H - c_L) + R_H^2 \right\} \\ &= R_H + \left[Q(p_H) + \delta Q(c_H) \right] \Delta^c \,. \end{aligned} \tag{46}$$

If the incentive compatibility constraint (46) binds and the participation constraint of the high-cost firm binds (so $R_H = 0$), then total discounted welfare is

$$\begin{aligned} W = \phi \left\{ w_L(p_L) + \delta w_L(c_L) - (1-\alpha)\Delta^c \left[Q(p_H) + \delta Q(c_H) \right] \right\} \\ +(1-\phi) \left\{ w_H(p_H) + \delta w_H(c_H) \right\} \;. \end{aligned} \tag{47}$$

Maximizing expression (47) with respect to the remaining choice variables, p_L and p_H, implies 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 (47) into the welfare achieved in the first period and the welfare achieved in the second period. Doing so gives:

$$\begin{aligned} W = \underbrace{\phi\{w_L(c_L) - (1-\alpha)\Delta^c Q(p_H)\} + (1-\phi) w_H(p_H)}_{\text{welfare from Baron-Myerson regime}} \\ + \delta \underbrace{\left[\phi\{w_L(c_L) - (1-\alpha)\Delta^c Q(c_H)\} + (1-\phi) w_H(c_H)\right]}_{\text{welfare from Loeb-Magat regime}} \,. \end{aligned} \tag{48}$$

Since the price p_H in expression (48) 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 in the Loeb-Magat regime, where 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. (Recall the discussion of the Loeb-Magat policy in section 2.1.1 above.) This second–period Loeb-Magat 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.[55] 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, both types of firm charge the same price $\tilde{p}$, say, in the first period, while 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 (48), the pooling regime results in lower welfare in the first period and higher welfare in the second. Much as in expression (48), total discounted welfare under this policy is

$$W = \underbrace{\phi\{w_L(\tilde{p}) - (1-\alpha)\Delta^c Q(\tilde{p})\} + (1-\phi)w_H(\tilde{p})}_{\text{welfare from pooling regime}}$$

$$+\,\delta\,\underbrace{\left[\phi\{w_L(c_L) - (1-\alpha)\Delta^c Q(p_H)\} + (1-\phi)w_H(p_H)\right]}_{\text{welfare from Baron-Myerson regime}}.$$

Whenever the discount factor δ is sufficiently large, 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.[56] When separation is not optimal, the precise details of the optimum are intricate. In rough terms, when the discount factor δ is small 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.[57]

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

[55] 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 section 10.6 of Laffont and Tirole (1993b).

[56] In fact, 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.

[57] See chapter 10 of Laffont and Tirole (1993b) for details of the solution. One technical issue is 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 page 390 of their book), but Bester and Strausz show that the contracts Laffont and Tirole consider include the optimal contracts.) For additional analysis regarding the design of contracts in the presence of adverse selection and renegotiation, see Rey and Salanié (1996), for example.

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.

2.3.3 Short-term Contracts: The Danger of Expropriation

Next consider the two-period setting of section 2.3.2 with one exception: the regulator cannot credibly commit to deliver specified second-period rents.[58] 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 rents. Unlike the renegotiation model, there are no long-term contracts that can defend the firm against this kind of 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 2.3.2. Suppose the regulator offers the two 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 fixed cost. Given that neither firm will receive any rent in the second period with 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

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

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 (49) is precisely the incentive compatibility constraint (46) for the low-cost firm in the setting with renegotiation. Assuming that incentive constraint (49) binds and the participation constraint for the high-cost firm binds, welfare is given by expressions (47) and (48). Natural candidates for optimal first-period prices are derived by maximizing this expression with respect to p_L and p_H, which are those 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.[59]

[58] This discussion is based on Laffont and Tirole (1988a) and chapter 9 of Laffont and Tirole (1993b). Freixas, Guesnerie, and Tirole (1985) explore a related model which considers only linear contracts.

[59] Laffont and Tirole call this the "take the money and run" strategy. This possibility is one of the chief differences between the no-commitment and the renegotiation scenarios. Under renegotiation, transfers and rents can be structured over time so that this is never a profitable strategy 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.

To determine when the incentive compatibility constraint for the high-cost firm binds, notice that when it 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

$$\begin{aligned} 0 &\geq Q(c_L)(c_L - c_H) - F + T_L \\ &= [Q(p_H) + \delta Q(c_H) - Q(c_L)]\,\Delta^c\,. \end{aligned} \tag{50}$$

When p_H is as specified in equation (3) in Proposition 1, expression (50) 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.[60] 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 (50) 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 (3) 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 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 that the low-cost and high-cost firm always set distinct prices. Instead, the regulator will prefer to induce the distinct types of the firm to implement the same price in the first period with positive probability.

These conclusions are summarized in Proposition 7.

Proposition 7 *In the two-period setting with no intertemporal regulatory commitment, the optimal regulatory policy has the following features:*
(i) When δ is sufficiently small that inequality (50) 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.

[60] When private information is distributed continuously (rather discretely as presumed in this chapter), it is never feasible (let alone optimal) to have a fully revealing first-period set of contracts. Since, with full separation, any firm obtains zero rent in the second period, it will always pay a firm 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 section 9.3 of Laffont and Tirole (1993b).

The pooling identified in property (iii) of Proposition 7 illustrates an important principle.[61] 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.[62] This insight 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 below.

Another important feature of the outcome with no commitment (and also with renegotiation) is that, at least when δ is sufficiently small that 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.) To see this, note first that the high-cost firm makes no rent whether the regulator's commitment powers are limited or unlimited, and so is indifferent between the two regimes. Without commitment, expression (49) reveals that the low-cost firm makes discounted rent $[Q(p_H) + \delta Q(c_H)]\,\Delta^c$. With commitment, however, Proposition 2.3.1 reveals that the corresponding rent is only $[Q(p_H) + \delta Q(p_H)]\,\Delta^c$. Thus, just as the possibility of regulatory capture turns out to harm the firm once equilibrium responses are accounted for (recall the discussion in section 2.2.2), the firm also suffers when the regulator is better able to credibly promise not to expropriate 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.

2.4 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 costs 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.[63]

[61] 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 abilities to achieve the ideal outcome in this dynamic context.

[62] See Sappington (1986).

[63] We have been unable to identify a treatment of the regulatory moral hazard problem that parallels exactly the problem that we analyze in this section. For recent related discussions of the moral hazard problem, see chapter 4 of Bolton and Dewatripont (2002) and chapters 4 and 5 of Laffont and Martimort (2002). For an analysis of optimal risk-sharing between consumers and the regulated firm in a full information framework, see Cowan (2002).

The simple moral hazard setting considered here parallels the framework of section 2.1.3 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 $D(\phi) \geq 0$ 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.[64] 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 therefore

$$U = \phi U_L + (1-\phi)U_H - D(\phi) \geq U^0 \ , \tag{51}$$

where expression (51) indicates that the firm must achieve expected utility of at least U^0 if it is to be induced to produce. The firm will only implement a strictly positive success probability ($\phi > 0$) if it is promised a higher utility in state L than in state H. The firm's optimal choice of 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, the firm's equilibrium level of effort, $\hat{\phi}(\Delta^U)$, satisfies:

$$D'(\hat{\phi}(\Delta^U)) \equiv \Delta^U \ . \tag{52}$$

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.[65] 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 (51) and the equilibrium effort condition $\phi = \hat{\phi}(\Delta^U)$ as defined by expression (52). Given the presumed separability in the firm's utility function, the participation constraint (51) will bind at the optimum. Therefore, we can re-state the regulator's problem as maximizing social surplus

$$W = \phi W_L(U_L) + (1-\phi)W_H(U_H) - D(\phi) \ , \tag{53}$$

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

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

where $W_i(U_i) \equiv V_i(U_i) + U_i$, subject to $\phi = \hat{\phi}(\Delta^U)$ and the participation constraint (51).

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 $R_i = \pi_i(p_i) + T_i$.[66]

Case 1: Risk-neutral firm when lump-sum transfers are used

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 $v_i(p_i) - T_i$ achievable for a given level of utility, is given by

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

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

$$W_i(U_i) \equiv w_i(p_i^*) \ , \tag{54}$$

and the size of total available surplus does not depend on how much rent/utility the firm is afforded.

Case 2: Risk-averse firm when lump-sum transfers are used

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

$$V_i(U_i) = w_i(p_i^*) - u^{-1}(U_i) \ , \tag{55}$$

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 firm utility and consumer surplus does not depend on the prevailing state. Consequently,

$$V_L'(U) \equiv V_H'(U) \ . \tag{56}$$

Case 3: Risk-neutral firm when no lump-sum transfers are used

When the firm is risk neutral, its utility is equal to its rent, as noted above. When no lump-sum transfers are employed $T_i \equiv 0$, and so $U_i = R_i = \pi_i(p_i)$. 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,

$$V_i(U_i) = v_i\left(\pi_i^{-1}(U_i)\right) \ . \tag{57}$$

[66] 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, 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, it follows that

$$V_i'(U_i) = \frac{-1}{1-\eta\left[\frac{p_i-c_i}{p_i}\right]} \quad , \tag{58}$$

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

Full information benchmark: First consider the case where the regulator can directly control effort ϕ, so that the effort selection constraint, $\phi = \hat{\phi}(\Delta^U)$, can be ignored. If λ is the Lagrange multiplier for the participation constraint (51) in this full information problem, the optimal choices for U_L and U_H satisfy

$$V_L'(U_L) = V_H'(U_H) = -(1+\lambda)\,. \tag{59}$$

Expression (59) 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 impose the constraint $\phi = \hat{\phi}(\Delta^U)$. In this second-best problem, if $\hat{\lambda}$ is the Lagrange multiplier associated with (51), then the first-order conditions for the choice of U_i are

$$V_L'(U_L) = -(1+\hat{\lambda}) - \frac{\hat{\phi}'}{\hat{\phi}}\Delta^V \;;\; V_H'(U_H) = -(1+\hat{\lambda}) + \frac{\hat{\phi}'}{1-\hat{\phi}}\Delta^V \;, \tag{60}$$

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 $\hat{\phi}' \approx 0$, expression (60) collapses to the full-information condition in (59), and so the full-information outcome is attained.[67]

In the ensuing sections we discuss the special cases of optimal regulation of a risk-neutral firm (case 1 in the above discussion) and a risk-averse firm (case 2). We defer discussion of the case of limited regulatory instruments (case 3) until section 3.3.

2.4.1 Regulation of a Risk-Neutral Firm

It is well known that when the firm is risk neutral, the full-information outcome is attainable. To see why, substitute (54) into expected welfare (53). Doing so reveals that the regulator's objective is to maximize

$$W = \phi w_L(p_L^*) + (1-\phi)w_H(p_H^*) - D(\phi) \tag{61}$$

subject to $\phi = \hat{\phi}(\Delta^U)$ and the participation constraint in (51). The regulator can structure the two utilities U_L and U_H to meet the firm's participation constraint (51) without affecting

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

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 the expression (61), 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 lump-sum transfers are used. The optimal outcomes for the firm and for consumers are:*

$$D'(\phi) = U_L - U_H = w_L(p_L^*) - w_H(p_H^*) \ ; \ V_L(U_L) = V_H(U_H) \, . \tag{62}$$

The conclusion in Proposition 8 parallels the conclusion in the Loeb-Magat model of regulation under adverse selection when distributional concerns are absent, discussed in section 2.1.1. In both cases, the firm is made the residual claimant for the social surplus and consumers are indifferent about which state of the world is realized. 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 subsidy of $T_i = v_i(p_i) - K$. 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 (61).

2.4.2 Regulation of a Risk-Averse Firm

When the relationship between firm utility and net consumer surplus is as specified in equation (55), conditions (56) and (59) together imply that if the regulator could directly control the firm's effort ϕ, the outcomes for consumers and the firm at the optimum would be:

$$U_L = U_H \ ; \ V_L(U_L) - V_H(U_H) = w_L(p_L^*) - w_H(p_H^*) \, . \tag{63}$$

In words, if the firm's effort could be controlled by other means, the risk-averse firm would be fully insured, so that it would receive the same utility (and rent) in the two states. Of course, full insurance leaves the firm with no incentive to achieve the desirable outcome. In contrast, a high-powered scheme provides strong effort incentives, but leaves the firm exposed to substantial risk.

The second-best policy is given by expression (60) 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.[68] 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 (60) implies that

$$U_L > U_H \ ; \ 0 < V_L(U_L) - V_H(U_H) < w_L(p_L^*) - w_H(p_H^*) \, . \tag{64}$$

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

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

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 (51) becomes

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

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

2.4.3 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, we now modify the model of section 2.4.1 to incorporate *ex post* participation constraint that the firm must receive rent $R_i \geq 0$ 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.[69] 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 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 again be important for the analysis. Therefore, as with the adverse selection analysis, suppose the regulator places the weight $\alpha \in [0, 1]$ on the firm's rents. In this case, much as in section 2.4.1 above, the regulator's objective is to choose R_L to maximize

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

(As before, it is optimal to set the full-information prices p_i^* and to use transfers to provide effort incentives.) Alternatively, since the incentive constraint is defined by the equality $R_L = D'(\phi)$, the regulator can be viewed as choosing ϕ to maximize

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

The solution to this problem has the first-order condition

$$D'(\phi) = w_L(p_L^*) - w_H(p_H^*) - (1 - \alpha)\phi D''(\phi) . \tag{65}$$

Comparing expression (65) with expression (62) the corresponding expression from the setting where there are no *ex post* participation constraints, it is apparent that these constraints result in less equilibrium effort. (Recall that $D'' > 0$.) Therefore, the introduction of a limited liability constraint lowers the power of the optimal incentive scheme. The lower 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

[69] We will assume the *ex ante* participation constraint does not bind in the ensuing analysis. See section 3.5 of Laffont and Martimort (2002) for further discussion of limited liability constraints.

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

In some respects this limited liability setting is similar to the case of risk aversion in section 2.4.2, because the full-information outcome is not feasible in either setting, and too little effort is supplied relative to the full-information outcome. Perhaps a closer parallel, however, is with the adverse selection analysis in section 2.1.1. The trade-off for the regulator is not between *insurance* and incentives, as it is in the model of moral hazard with a risk-averse firm, but between *rent extraction* and incentives.

2.4.4 Repeated Moral Hazard

The extension of the static analysis of regulation under moral hazard to a dynamic setting is possible using recent techniques developed for the general principal-agent problem. However, a full treatment would be beyond the scope of this chapter, especially since several of the insights parallel those derived in the adverse selection setting of section 2.3. There are three main additional features that are introduced when the moral hazard model is repeated over time.[70] First, the firm could 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 run 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 could choose to respond to a positive outcome in the current period by reducing effort to some extent in the future. Consequently, the regulator's optimal inter-temporal policy, and the firm's profit-maximizing response to the policy, can be complicated.[71] 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 current period's outcomes.

The dynamic moral hazard problem is discussed further in section 3.2.3 below, when we analyze the optimal frequency of regulatory review.

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

[70] See chapter 11 of Bolton and Dewatripont (2002), which also emphasizes the effects of limited commitment on the part of the principal. See also the analyses of renegotiation by Fudenberg and Tirole (1990), Chiappori, Macho, Rey, and Salanié (1994), Ma (1994), and Matthews (2001). Section 8.2 of Laffont and Martimort (2002) analyzes the two-period model with full commitment. Also see Radner (1981, 1985) for early work on the repeated moral hazard problem.

[71] See Rogerson (1985). Holmstrom and Milgrom (1987) show that the optimal inter-temporal incentive scheme is linear in the agent's total production in a particular continuous time framework.

The analysis in this section has focused on the design of optimal regulatory policy when there is a single monopoly supplier of regulated services.[72,73] 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 analyses of optimal regulatory policy. This approach models formally the information asymmetry between the regulator and the firm and then determines precisely how the regulator optimally pursues his goals in the presence of this asymmetry. However, in practice: (i) all relevant information asymmetries can be difficult to characterize precisely; (ii) a complete specification of all relevant constraints on the regulator and firm can be difficult to formulate; (iii) some of the instruments that are important in optimal reward structures (such as lump-sum transfers) are not always available; and (iv) 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 typically are incapable of capturing the full richness of the settings in which actual regulatory policies are implemented.[74]

This fact 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 discussed under four headings: (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. Although these headings incorporate substantial overlap, the four categories are useful for pedagogical purposes.[75]

[72] The analysis in this section also has also 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 (1993, chapter 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.

[73] 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 next in section 3.

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

[75] 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

To begin, it may be helpful to assess how two of the most familiar regulatory policies compare on these four dimensions. Table 1 provides a highly stylized interpretation of how price cap and rate-of-return regulation differ along these dimensions.

Table 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	Yes	No

Table 1 reflects the idea that at least under an extreme form 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 regime has expired). By contrast, under an extreme form of rate-of-return regulation: (i) the regulator sets prices, and affords the firm little or no discretion in altering these prices; (ii) prices are adjusted as necessary to ensure that the realized rate of return on investment does not deviate 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.[76]

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, namely, *average revenue* regulation and *tariff basket* regulation.

Despite the potential merits of delegating some pricing flexibility to the regulated firm, there are reasons why regulators might wish to limit the firm's pricing discretion. One reason

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.

[76] For more detailed discussions of the key differences between price cap regulation and rate-of-return regulation, see, for example, Acton and Vogelsang (1989), Hillman and Braeutigam (1989), Braeutigam and Panzar (1993), Liston (1993), Armstrong, Cowan, and Vickers (1994), Blackmon (1994), Mansell and Church (1995), Sappington (1994, 2002) and Sappington and Weisman (1996a).

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 undoing the cross subsidies that regulators have imposed historically to promote social goals such as universal service.

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.[77] We assume that no transfer payments to the firm are permitted, and the firm's tariff must be designed to cover its costs. Suppose further that only linear tariffs are used.[78] As in section 2, the regulator seeks to maximize a weighted average of expected consumer surplus and profit, where $\alpha \leq 1$ 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 $\mathbf{p} = (p_1, ..., p_n)$. 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.

In this setting some pricing flexibility is always advantageous. To see why, suppose the regulator instructs the firm to offer the fixed price vector $\mathbf{p}^0 = (p_1^0, ..., p_n^0)$. 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 just as well off as they were with $\mathbf{p}^0$, so that the firm can choose any price vector

$$\mathbf{p} \in \mathcal{P} = \left\{\mathbf{p} \mid v(\mathbf{p}) \geq v(\mathbf{p}^0)\right\} . \tag{66}$$

Figure 2 illustrates $\mathcal{P}$, the set of prices the firm can offer under this form of regulation, for the case where the firm provides two products. $\mathcal{P}$ is the shaded region comprised of those prices that lie below the contour that leaves consumers indifferent to the price vector $\mathbf{p}^0$.

By construction, this regulatory policy ensures that consumers in aggregate are no worse off than they are under the fixed pricing policy $\mathbf{p}^0$.[79] 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, total welfare is sure to increase when the firm is granted pricing flexibility in this way.[80]

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

[78] However, this does not rule out two-part tariffs. If two-part tariffs are offered, "access" should be defined as a separate product, and the fixed part of the two-part tariff can be viewed as the price of access.

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

[80] Notice that the profit-maximizing prices for the firm operating under this constraint are closely related to Ramsey prices: profits are maximized subject to a consumer surplus constraint or, equivalently, consumer

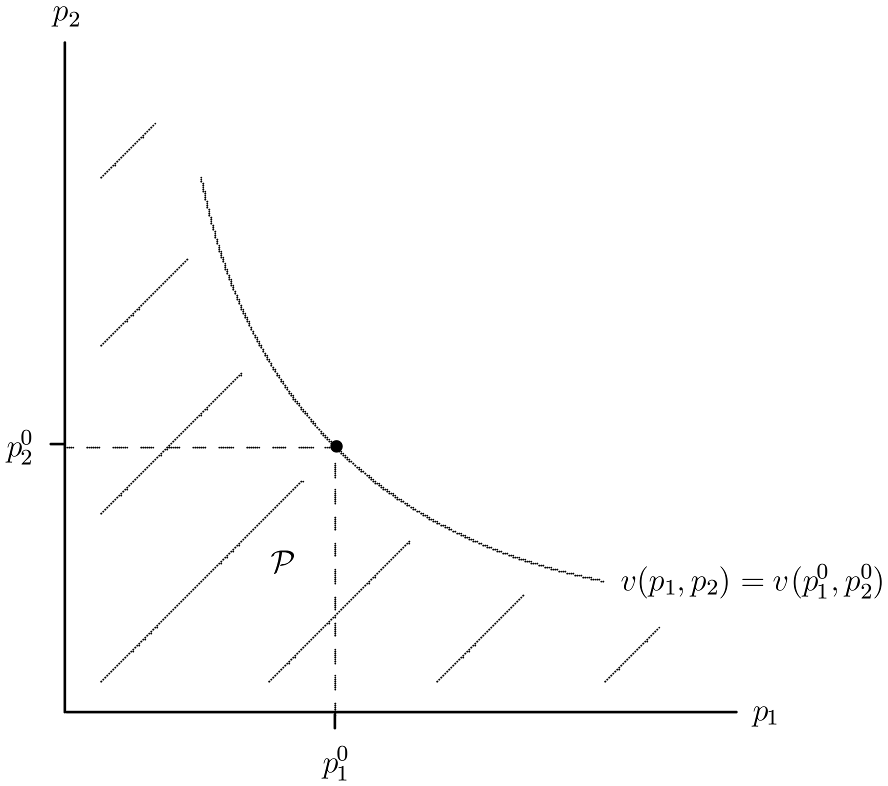

Figure 2: The Benefits of Tariff Flexibility with Known Demands

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 unit costs $\mathbf{c} = \{c_1, ..., c_n\}$ for its n products. Then 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 prices 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 is closely related to the analysis in section 2.1.2 of the optimal regulation

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.

(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 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 will typically be better informed than the regulator about both its demand and its cost structure, and the regulator will often be unaware of the precise form 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 principles outlined above can inform the choice of the degree of pricing flexibility afforded the firm.

The next section discusses the performance of two common methods for granting the firm some pricing flexibility.

3.1.2 Forms of Price Flexibility

The merits of affording the regulated firm some pricing flexibility will 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.[81] Suppose the demand function for the i^{th} product 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, the following expression from consumer demand theory is useful. For any pair of price vectors $\mathbf{p}_1$ and $\mathbf{p}_2$ the following inequality holds:[82]

$$v(\mathbf{p}^2) \geq v(\mathbf{p}^1) - \sum_{i=1}^{n}(p_i^2 - p_i^1)Q_i(\mathbf{p}^1) \; . \tag{67}$$

Expression (67) 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 expression follows from the convexity of the consumer surplus function.

Average Revenue Regulation: In its simplest (static) 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 be in the set

$$\mathbf{p} \in \mathcal{P}^{AR} = \left\{ \mathbf{p} \mid \frac{\sum_{i=1}^{n} p_i Q_i(\mathbf{p})}{\sum_{i=1}^{n} Q_i(\mathbf{p})} \leq \bar{p} \right\} \; . \tag{68}$$

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

[82] The right-hand side of this expression 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.

The term to the left of the inequality in expression (68) is average revenue: total revenue divided by total output.[83] 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 (68) exactly, then inequality (67) implies that $v(\mathbf{p}^1) \leq v(\mathbf{p}^2)$. Therefore, this form of average revenue regulation will leave consumers worse off compared to a uniform pricing regime, regardless of the firm's chosen prices.[84] 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 3 for the case where the firm offers two products. Here the boundary of the set $\mathcal{P}^{AR}$ in (68) lies inside the set of price vectors that make consumers worse off than they are with the uniform price vector $(\bar{p}, \bar{p})$.

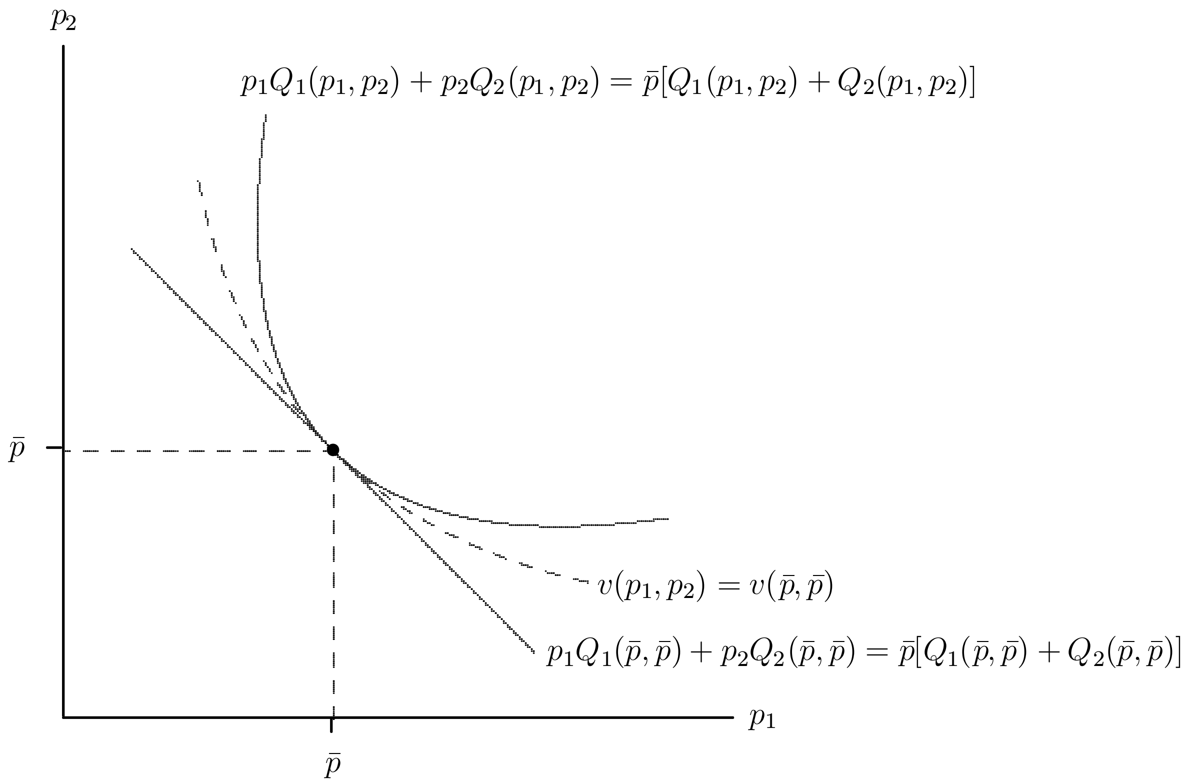

Figure 3: Tariff Basket and Average Revenue Regulation

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

Proposition 9 *(i) Consumer surplus is lower under binding average revenue regulation when the firm is permitted to set any prices that satisfy inequality (68) rather than being required to set each price at $\bar{p}$.*

[83] 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.

[84] Armstrong, Cowan, and Vickers (1995) show that, for similar reasons, allowing nonlinear 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.

(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 (68) 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, and 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 (68), 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 $\bar{p}$.[85]

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 (68) is

$$p + \frac{A}{Q(p)} \leq \bar{p} \ . \tag{69}$$

Inequality (69) 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 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.[86] The profit-maximizing behavior of the firm under the average revenue constraint in inequality (69) 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 (69) will bind, and so the firm's

[85] See Bradley and Price (1988), Law (1995), and Cowan (1997b). 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.

[86] See Sappington and Sibley (1992) and Cowan (1997a) 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).

profit (per consumer) 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.[87] Consequently, average revenue regulation in this setting induces a very 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 is allowed a margin $\bar{p} - c$ per unit of its output, and so it has an incentive to expand output inefficiently.[88]

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 would 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:

$$\mathbf{p} \in \mathcal{P}^{TB} = \left\{ \mathbf{p} \mid \sum_{i=1}^{n} p_i Q_i(\mathbf{p}^0) \leq \sum_{i=1}^{n} p_i^0 Q_i(\mathbf{p}^0) \right\} . \tag{70}$$

Under this form of tariff basket regulation, the weights that are implicitly employed to calculate the firm's average price are fixed from the firm's perspective, 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 $\mathbf{p}^0$. (This follows from formula (67) 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 (70).) This form of tariff basket regulation parallels the regulatory policy specified in expression (66). In particular, the set of prices in (70) lies inside the set (66) which, by construction, is the set of prices that make consumers better off than with $\mathbf{p}^0$.

This finding is illustrated in Figure 3 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 (70) 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 (70). Since the firm will also be better off with the flexibility permitted in constraint (70), 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 has the price

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

[88] This conlusion is similar to Averch and Johnson (1962)'s 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.

p^0. In this case, constraint (70) becomes

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

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

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

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 more total 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 $\mathbf{p}^0$. 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, though, outputs in the previous period can serve 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.[89] 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:

1. *Laissez-faire*: Here the incumbent can pursue any pricing policy in the two markets it chooses.

2. *Ban on price discrimination*: Here the incumbent can choose 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 in non-competitive markets.) 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 an immediate threat in that market.

3. *Separate price caps*: Here the incumbent faces only an upper limit on the price it can charge in each market, and so can price below the cap in the market where entry occurs. Importantly, the two price caps are not linked, in the sense that the price set in one market has no effect on the price the firm can charge in the other market.

4. *Average price cap*: Here the incumbent operates under an average price cap for the two markets. Therefore, if the incumbent lowers its price in one market in response to entry, it can then *raise* its price in the other market. Thus, in contrast to the ban on price discrimination, here there is an inverse relationship between feasible prices in the two markets.

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

Here, regimes 1 and 2 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 incumbent responses to entry. 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.[90] 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. In particular, the benefits 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—discussed in the monopoly setting in sections 3.1.1 and 3.1.2—are less clear-cut 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. First, section 3.2.1 considers various kinds of dynamic average price regulation. In particular, no transfers from the regulator to the firm are permitted, and so the main feature of interest is how the current allowed set of prices depends on the history of regulation (e.g., the prices chosen by the firm in the past, or the observed profits of the firm). Second, section 3.2.2 extends the analysis to allow the regulator to make transfers to the firm. Third, section 3.2.3 examines how frequently the regulator should realign the firm's prices to match its observed costs. Finally, section 3.2.4 discusses the effect of (exogenous) technological change on the inter-temporal pattern of prices.

[90] Armstrong and Vickers (1993) model this interaction as a Stackelberg price game, in which the entrant maximizes its profit, taking the incumbent's (post-entry) price as given.

3.2.1 Non-Bayesian Price Adjustment Mechanisms: No Transfers

First consider the natural dynamic extension of the tariff basket form of average price regulation analyzed in section 3.1.2. In this dynamic extension, the weights employed in the current price cap reflect the previous period's outputs.[91] Call the initial period in this dynamic setting period '0', and label subsequent periods $t = 1, 2, ...$ Let $\mathbf{p}^t = (p_1^t, ..., p_n^t)$ denote the vector of prices the firm charges for its n regulated products in period t. Let $\mathbf{q}^t = (q_1^t, ..., q_n^t)$ denote the corresponding vector of output levels, 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 the previous period, the firm can choose any price vector $\mathbf{p}^t$ in the current period satisfying

$$\mathbf{p}^t \in \mathcal{P}^t = \left\{ \mathbf{p}^t \mid \sum_{i=1}^{n} p_i^t q_i^{t-1} \leq \sum_{i=1}^{n} p_i^{t-1} q_i^{t-1} \right\} . \tag{71}$$

(We discuss below how the initial price vector $\mathbf{p}^0$ might be determined, but for now $\mathbf{p}^0$ is taken to be specified exogenously.) 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 (71) can be written as

$$\mathbf{p}^t \in \mathcal{P}^t = \left\{ \mathbf{p}^t \mid \sum_{i=1}^{n} \frac{R_i^{t-1}}{R^{t-1}} \left[\frac{p_i^t - p_i^{t-1}}{p_i^{t-1}} \right] \leq 0 \right\} , \tag{72}$$

where $R_i^{t-1} = p_i^{t-1} q_i^{t-1}$ is the revenue generated by the ith product in period $t-1$, and R^{t-1} is total revenue from the n products in period $t-1$. Constraint (72) 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.

Figure 4 illustrates how this form of dynamic average price regulation evolves. For the reasons explained in section 3.1.2, any price vector in the set defined by (71) generates at least as much consumer surplus as the previous period, so $v(\mathbf{p}^t) \geq v(\mathbf{p}^{t-1})$. 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 could choose the same vector $\mathbf{p}^0$ in each period if it wished, the firm must be better off as well. This dynamic process converges and the steady state price vector will have the Ramsey form: profit is maximized subject to a consumer surplus constraint.[92] (However, as in section 3.1.1, the long-run prices are not precise Ramsey prices since the firm's rent will not in general be zero.)

Regarding the initial price vector $\mathbf{p}^0$, the regulator might choose these prices to ensure that the firm makes only small rents in the long term and that total discounted expected welfare is maximized. Such a choice would require a substantial amount of information, however. Alternatively, $\mathbf{p}^0$ might be set by the firm without constraint—so that the firm

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

[92] 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 prices, $\mathbf{p}^{t-1}$. From Figure 4, this implies that the firm's iso-profit contour is tangent to the iso-consumer surplus contour.

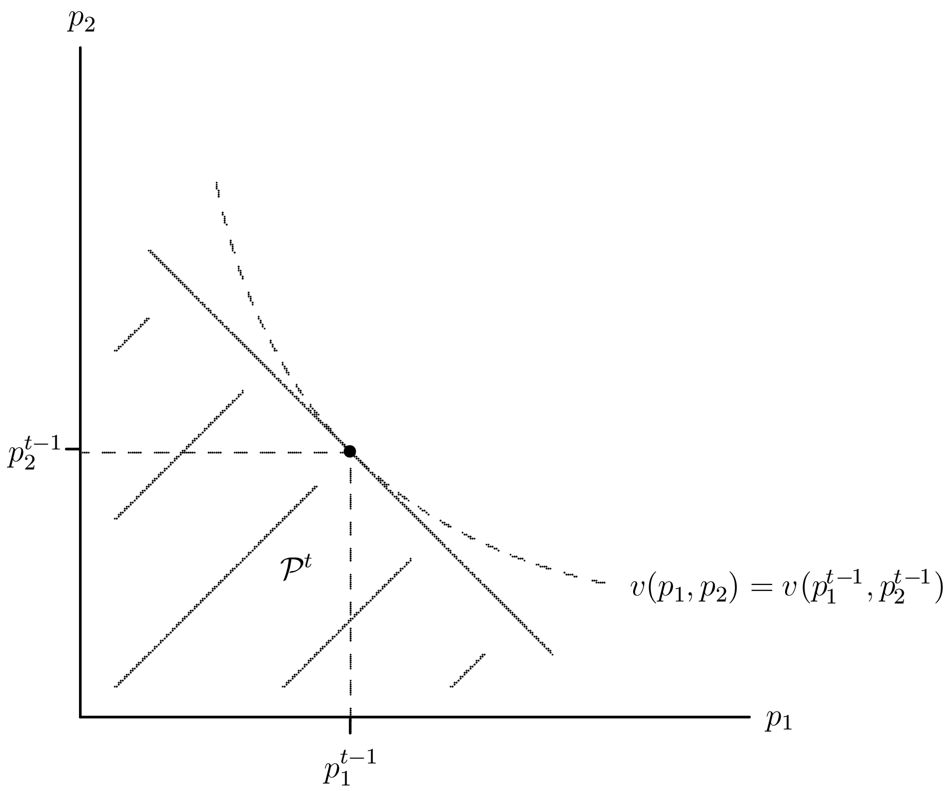

Figure 4: Dynamic Tariff Basket Regulation

is initially unregulated—but where the firm expects subsequently to be controlled by the regulatory mechanism (71). In this setting, the firm will set its initial prices strategically in order to affect the weights in future constraints. For instance, the firm 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 might be to reduce aggregate welfare below the level achieved in the absence of any regulation.[93]

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.[94]

Although this form of dynamic regulation leads to an increase in consumer surplus in every period, it does not ensure a particularly high level of 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

[93]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.

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

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.[95] Formally, the constraint (72) is then modified to:

$$\mathbf{p}^t \in \mathcal{P}^t = \left\{ \mathbf{p}^t \mid \sum_{i=1}^{n} \frac{R_i^{t-1}}{R^{t-1}} \left[\frac{p_i^t - p_i^{t-1}}{p_i^{t-1}} \right] \leq -X \right\} . \tag{73}$$

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, albeit in a restrictive model, consider the following policy, referred to as the *VF* mechanism.[96] The mechanism 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. The firm's (observable) expenditures in year t are $E^t \geq C(\mathbf{q}^t)$, where $C(\mathbf{q}^t)$ is the minimum possible cost of producing output vector $\mathbf{q}^t$.[97] Then the *VF* mechanism permits the firm in period t to select any vector of prices that lie in the set

$$\mathbf{p}^t \in \mathcal{P}^t = \left\{ \mathbf{p} \mid \sum_{i=1}^{n} p_i q_i^{t-1} \leq E^{t-1} \right\} . \tag{74}$$

The *VF* mechanism differs from the regulatory regime reflected in expression (71) in that last period's expenditure replaces last period's revenue as the cap on the current level of calculated revenue. If we let $\Pi^t = \sum_{i=1}^{n} p_i^t q_i^t - E^t$ denote the firm's (observable) profits in period t, constraint (74) can be re-written as

$$\mathbf{p}^t \in \mathcal{P}^t = \left\{ \mathbf{p} \mid \sum_{i=1}^{n} p_i q_i^{t-1} \leq \sum_{i=1}^{n} p_i^{t-1} q_i^{t-1} - \Pi^{t-1} \right\} .$$

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 decrease sufficiently to eliminate the observed profit of the firm in the previous period (and not simply decrease, as in expression (71)). In particular, expression (67) shows that $v(\mathbf{p}^t) \geq v(\mathbf{p}^{t-1}) + \Pi^{t-1}$, and so "excess profits" in one period are (more than) transferred to consumers in the next period. Notice that the regulator only needs to observe the firm's realized sales and expenditures in order to implement the *VF* 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 *VF* mechanism can induce desirable outcomes under certain stringent conditions. In particular, the *VF*

[95] We will discuss other aspects of this issue in section 3.2.4 below.

[96] *VF* denotes Vogelsang and Finsinger (1979), the authors who proposed this regulatory mechanism.

[97] For simplicity, we abstract from intertemporal cost effects, so that all costs of producing output $\mathbf{q}^t$ are incurred in period t.

mechanism can sometimes eventually induce exact Ramsey prices (i.e., the prices that maximize consumer surplus while securing non-negative rent for the firm). This conclusion is summarized in Proposition 10.

Proposition 10 *Suppose that demand and cost functions do not change over time and that the firm's technology exhibits decreasing ray average cost.*[98] *Suppose further that the firm maximizes profit myopically each period. Then the VF mechanism induces the firm to set prices that converge to the Ramsey prices.*

The conditions under which the *VF* mechanism 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 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 rule (74) 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 $c^t \geq c$, where c is its "innate", or minimum, unit cost. Thus, any choice $c^t > c$ constitutes "pure waste". (Note that this inflated cost is actually incurred by the firm.) The firm discounts future profits 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 (somehow) 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 set a higher cost $c_H \geq c$ in period 0, and then implement the minimum cost c in every period thereafter. With this particular strategy, the firm's discounted profit is

$$Q(p^0)(p^0 - c_H) + \delta Q(c_H)(c_H - c) \ .$$

This expression is increasing in c_H at $c_H = c$ when $\delta Q(c) > Q(p^0)$, in which case the firm is able to increase its profit by inflating its cost in period 0. Therefore, whenever the discount factor is high enough—so that the firm cares sufficiently about future profit—the firm will find it profitable to inflate its costs.[99]

[98] The cost function $C(q)$ exhibits decreasing ray average cost if $rC(q) \geq C(rq)$ for all $r \geq 1$. Decreasing ray average costs ensure the firm can continue to secure non-negative profit under the VF mechanism as prices decline and outputs increase.

[99] This discussion assumes the firm can *costlessly* reduce its costs to the efficient level. If the firm must incur higher (unobserved) fixed cost in order to reduce marginal costs, this effect could be amplified. Sappington (1980) shows that because of the pure waste it can induce, the *VF* mechanism 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 *VF* mechanism 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 *VF* mechanism imposes on prices.

The two dynamic price regulation mechanisms reviewed in this section affect the prices of the multiproduct firm along two dimensions: the pattern of relative prices, and the average price level. The tariff-basket adjustment mechanism reflected in constraint (71) 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 rents indefinitely. The *VF* mechanism attempts to overcome this drawback. The *VF* mechanism delivers a desirable equilibrium pattern of relative prices. It also eliminates 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 price levels.

3.2.2 Non-Bayesian Price Adjustment Mechanisms: Transfers

Incentives for pure waste (or related distortions) do not arise under another regulatory mechanism that requires little knowledge of industry conditions to design and implement. This mechanism, called the FV subsidy mechanism, requires monetary transfers from the regulator to the firm, but can eventually induce the firm to set prices equal to marginal production costs.[100] The FV subsidy mechanism operates as follows when, for expositional simplicity, the regulated firm produces only a single product. Each period, the regulated firm is permitted to set any price (p^t) it desires for its product. The firm retains all of the profit it generates each period. Actual profit in period t is denoted Π^t and is observed by the regulator. Given its performance in the previous period, the firm also receives the following subsidy in period t:

$$S^t = q^{t-1}[p^{t-1} - p^t] - \Pi^{t-1} . \tag{75}$$

This subsidy is the difference between: (1) an approximation to the increment in consumer surplus derived from any price reduction the firm implements in period t; and (2) the firm's profit in period $t - 1$.

The FV subsidy mechanism induces the regulated firm to maximize an approximation to the increment in total surplus it generates each period. Consequently, the firm maximizes the present discounted value of its net payoffs by gradually reducing price toward marginal cost. Furthermore, pure waste does not relax a binding constraint on prices (as it can under the *VF* mechanism), and so is never optimal for the firm. Thus, when demand and cost functions do not change over time, the FV subsidy mechanism ultimately achieves the outcome a welfare-maximizing regulator would implement if he shared the firm's private knowledge of its environment. These observations are summarized in Proposition 11.

[100] As originally proposed and analyzed in Finsinger and Vogelsang (1981, 1982), the FV subsidy mechanism was designed to motivate public enterprises. The ensuing discussion adapts the original FV subsidy mechanism to apply to profit-maximizing regulated firms.

Proposition 11 *When it operates under the FV subsidy mechanism in a stationary environment, the regulated firm never undertakes pure waste. Furthermore, it sets prices that converge to the firm's marginal cost of production, and the firm's rent converges to zero.*[101]

Although the *FV* subsidy mechanism ultimately secures the welfare optimum in a stationary environment, it has at least three important drawbacks. First, the mechanism will not necessarily ensure the welfare optimum when demand and cost functions change over time. Rising costs or declining demand could even bankrupt a firm that operates under the *FV* subsidy mechanism. Second, the mechanism generally provides inadequate incentives for the firm to devote unobserved resources (such as managerial diligence and effort) to reduce operating costs.[102] Third, even in a stationary environment, convergence to the welfare optimum may be slow.[103]

If the regulator is perfectly informed about the demand curve facing the regulated firm, then the last of these drawbacks can be mitigated substantially. In fact, convergence to the full-information outcome is achieved in a single period. Here, the regulator awards the firm a subsidy each period equal to the actual (not approximate) increment in consumer surplus derived from its pricing decisions, less historic profit. Formally, this subsidy in period t is modify from expression (75) to be:

$$S^t = v(p^t) - v(p^{t-1}) - \Pi^{t-1}, \tag{76}$$

where $v(\cdot)$ is the consumer surplus function associated with the (known) demand function $Q(\cdot)$, and Π^{t-1} is again observed profit in period $t-1$. Call this subsidy mechanism the incremental surplus subsidy (*ISS*) mechanism.[104]

To illustrate the workings of this dynamic mechanism in the simplest case, suppose there is an exogenous profit function $\pi(p^t)$, the precise form of which is not known to the regulator. (However, as before, the actual profits $\Pi^t = \pi(p^t)$ are observed.) Let $w(p) \equiv v(p) + \pi(p)$ be total per-period surplus generated when price p is set p. Also, let $R(p^{t-1})$ be the discounted maximized rent of the firm in period t under this *ISS* scheme, given that the price chosen in

[101] Finsinger and Vogelsang (1981, 1982, 1985) prove this convergence result. Vogelsang (1988) proves that pure waste will not occur.

[102] Vogelsang (1983) and Gravelle (1985) show that incentives to reduce operating costs are enhanced if the *FV* subsidy mechanism is modified to deliver to the firm each year an approximation to the sum of all increments in surplus generated in all preceding years. By aggregating incremental surplus gains in this manner, the firm is effectively subsidized by the full amount of the surplus derived from its activities, and so finds it profitable to deliver effort to increase the surplus. Because it awards substantial ongoing subsidies to the firm, however, this mechanism can raise distributional concerns.

[103] However, the *FV* mechanism avoids another potential drawback to mechanisms such as that analyzed by Tam (1981) that deliver greater rewards to the firm as the price it charges declines. This drawback is cycling, whereby the firm lowers and raises prices repeatedly over time. Finsinger and Vogelsang (1985) show that the mechanism proposed by Tam (1981) can induce cycling. See Vogelsang (1988) for a related observation. Laffont and Tirole (1993, pp. 142–143) suggest extensions of Tam's mechanism that perform well if the regulator can predict the level of demand that will prevail at socially optimal price levels. Of course, such predictions will be problematic in practice if the regulator's knowledge of demand and cost functions is truly limited.

[104] This is the name given to the mechanism by its authors, Sappington and Sibley (1988).

period $t-1$ was p^{t-1}. Dynamic programming implies this maximum rent can be written as:

$$\begin{aligned} R(p^{t-1}) &= \max_{p^t} : \left\{\pi(p^t) + S^t + \delta R(p^t)\right\} \\ &= \max_{p^t} : \left\{w(p^t) + \delta R(p^t)\right\} - w(p^{t-1}) , \end{aligned} \tag{77}$$

where expression (76) is employed to replace the period t subsidy in expression (77). Letting

$$K = \max_{p^t} : \left\{w(p^t) + \delta R(p^t)\right\} , \tag{78}$$

Expression (77) implies that $R(p) = K - w(p)$. Therefore, from (77), p^t is chosen to maximize $w(p) + \delta(K - w(p))$. Consequently, if $\delta < 1$ the price p^t is chosen to maximize $w(\cdot)$, and so the firm will implement marginal-cost pricing (denoted p^* say) in every period, starting with period 1. Expression (78) then implies that $K = w(p^*)$. Consequently, starting from an initial price p^0, the firm earns discounted rent in period 1 equal to the gain in total welfare from marginal-cost pricing in the first period, so[105]

$$R(p^0) = w(p^*) - w(p^0) . \tag{79}$$

Note that, in contrast to the *VF* mechanism, the *ISS* scheme does not provide incentives for the firm to distort its (observed) profits deliberately. 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 simply take the form of "pure waste", so that $\Pi = \pi(p) - e$ for some "true" profit function π, or it could be a choice variable that reflects, say, the firm's mix of fixed and marginal costs, or perhaps the quality of the firm's product.) Write $W(p, e) \equiv v(p) + \Pi(p, e)$ for the welfare function. Then if the previous period's choices were p^{t-1} and e^{t-1}, the rent of the firm under the *ISS* scheme is modified from (77) to be

$$R(p^{t-1}, e^{t-1}) = \max_{p^t, e^t} : \left\{W(p^t, e^t) + \delta R(p^t, e^t)\right\} - W(p^{t-1}, e^{t-1}) .$$

Like equation (77), this equation has the solution

$$R(p, e) = W(p^*, e^*) - W(p, e) ,$$

where p^* and e^* maximize $W(p, e)$. Therefore, from period 1 onwards, marginal-cost prices are set and the welfare-maximizing technology parameter e is implemented. Viewed from the perspective of the initial period 0, the firm's discounted rent is

$$\Pi(p^0, e^0) + \delta\left[W(p^*, e^*) - W(p^0, e^0)\right] .$$

Therefore, for a given initial price p^0, if the firm were free to choose the initial technology parameter e^0, and did so anticipating that the *ISS* scheme would be implemented from period 1 onwards, it would choose e^0 to maximize $(1-\delta)\pi(p^0, e)$. Therefore, there is no incentive for pure waste, even in the initial period.

In sum:[106]

[105] Given that the firm makes profit $\pi(p^0)$ in the initial period, its total discounted rent from the perspective of period 0 is $\pi(p^0) + \delta\left[w(p^*) - w(p^0)\right]$.

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

Proposition 12 *In a stationary environment the ISS mechanism 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.*

Although it may offer some improvements over the *FV* mechanism, the *ISS* mechanism has at least three main drawbacks. First, it can impose financial hardship on the firm if its costs rise over time.[107] Second, just as with the related Loeb and Magat (1979) mechanism discussed in section 2.1.1, the high subsidy payments that the ISS mechanism requires are socially costly when the regulator prefers consumer surplus to rent.[108] Third, although it avoids pure waste, the ISS mechanism 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, or the lower managerial effort required to produce at inefficiently high cost, for example.[109]

To understand why the regulated firm may undertake abuse under the *ISS* mechanism, consider a case where the regulator can observe some, but not all, components of the firm's costs. Specifically, suppose unit costs c are observed, while fixed costs (which represent the managerial effort associated with producing with unit cost c) F are not observed. Further suppose these costs are related by $F = F(c)$, where $F'(c) < 0$. Suppose the *ISS* transfer in period t, as in (76), takes the form

$$S^t = v(p^t) - v(p^{t-1}) - q^{t-1}(p^{t-1} - c^{t-1}) \,. \tag{80}$$

In this case the firm's rent in period t when its marginal cost in period $t-1$ is c^{t-1} and it sets the price p^{t-1} in period $t-1$ is:

$$R(p^{t-1}, c^{t-1}) = \max_{p^t, c^t} : \left\{ W(p^t, c^t) - F(c^t) + \delta R(p^t, c^t) \right\} - W(p^{t-1}, c^{t-1}) \,,$$

where $W(p, c) \equiv v(p) + Q(p)(p - c)$ is total welfare excluding the fixed costs $F(c)$. Then, as before, the solution takes the form $R(p, c) = K - W(p, c)$. However, in this case, p^t and c^t are chosen by the firm to maximize $(1-\delta)W(p, c) - F(c)$. Therefore, price p is set equal to realized cost c, but c is set at an inefficiently high level.[110] This is because the firm does not retain the full benefit of a unit cost reduction, since any profit generated in one period is fully usurped in the next period. (Notice from equation (80) that if, by incurring high fixed costs, the firm achieves a low marginal cost c^t in one period, it will receive a lower subsidy S^{t+1} in the subsequent period. Consequently, the firm will not appropriate the full benefits of its unobserved cost-reducing activity.)

In this sense, the *ISS* mechanism resembles cost-plus regulation, albeit with a one-period lag. The next section considers the issue of "regulatory lag" in more detail.

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

[108] Lyon (1996) presents simulations which suggest that once subsidy costs are accounted for, the *VF* mechanism (modified as Hagerman (1990) suggests to eliminate incentives for pure waste) often generates higher levels of welfare than the *ISS* mechanism.

[109] Sappington and Sibley (1993) show that the ISS 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 ISS mechanism.

[110] This distortion parallels the optimal distortion induced in the setting of Proposition 3 above.

3.2.3 Frequency of Regulatory Review

Even when regulatory regimes do not explicitly link prices to realized costs, they can implement partial cost-plus regulation through their updating procedures.[111] To illustrate, suppose that the authorized rate at which prices can rise (i.e., the X factor) in a price cap regulation regime is updated periodically to eliminate the firm's expected future rents. Also suppose that expectations about future profit are based in part upon the firm's current realized revenues and costs.[112] 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 for instance by efficiency gains—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 extranormal profit. On the other hand, an infrequent revision of the regime could mean that prices deviate from costs for long periods, which reduces allocative efficiency. The optimal choice of "regulatory lag" trades off these two opposing effects.[113] 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, so that these costs evolve according to some exogenous and possibly uncertain process, then frequent regulatory reviews are optimal. Since there is no need to give incentives for cost reduction in this case, the only concern is to achieve allocative efficiency, which can be accomplished through frequent reviews that set prices to match realized costs.

- If consumer demand is inelastic, so there is no deadweight welfare loss when prices depart from costs, then 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

[111] Explicit linkage of prices to costs is discussed in section 3.3.

[112] 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.

[113] 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 2.3.) Isaac (1991) points out that rate shock (substantial, rapid price changes) may occur if prices are revised to reflect realized operating costs only infrequently.

between review of price cap regimes generally will depend in a complicated manner upon the specifics of the regulatory environment.

The *VF* mechanism discussed in section 3.2.1 can be viewed as a regulatory regime with frequent regulatory reviews. Under the *VF* mechanism, the firm's prices in each period are required to fall to a level that just covers 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 the responsiveness of prices to realized costs in a dynamic setting. This issue is explored further in a static context in section 3.3.

3.2.4 Choice of 'X' in Price Cap Regulation

Recall from the discussion of expression (73) 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.[114]

Further suppose that investment, production and consumption all take place in periods $t = 0, 1, ...$ Let p_t denote the (linear) 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, there are no intertemporal linkages in demand. 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

$$K_{t+1} = (1 - d)K_t + \frac{I_{t+1}}{\beta_{t+1}} \ . \tag{81}$$

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 $K_t, K_{t+1}, ..., I_t, I_{t+1}, ...$ satisfying expression (81). Then if K_t is increased by 1, all subsequent values for K and I are unchanged if next period's investment I_{t+1} is reduced so as to the keep the right-hand side of expression (81) constant, i.e., if I_{t+1} is reduced by $(1 - d)\beta_{t+1}$.[115] If the interest rate is r, so that the firm's discount factor is

[114] The analysis in this section is based on section 4.4.1.3 of Laffont and Tirole (2000) and section 2.7 of Armstrong (2002). See also Kwoka (1991, 1993), section 6.3 of Armstrong, Cowan, and Vickers (1994), and Bernstein and Sappington (1999) for further discussions.

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

$\delta = \frac{1}{1+r}$, then the net cost of this modification to the investment plan is

$$C_t = \beta_t - \frac{1-d}{1+r}\beta_{t+1} \; . \tag{82}$$

Expression (82) specifies the marginal cost of obtaining a marginal 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, the above formula reduces to[116]

$$C_t = \beta_t \left(1 - \frac{(1-d)(1-\gamma)}{1+r}\right) \; . \tag{83}$$

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

Suppose that it costs the firm an amount c_t to convert a unit of capacity into a unit of the final product. Then total discounted welfare, measured as the sum of consumer surplus and profit, is

$$W = \sum_t \frac{1}{(1+r)^t} \left\{ v_t(p_t) + Q_t(p_t)(p_t - c_t) - I_t \right\} \; , \tag{84}$$

where $K_t \equiv Q_t$ and this capital stock evolves according to expression (81). Notice that expression (81) implies

$$I_t = \beta_t \left[Q_t - (1-d)Q_{t-1} \right] \; . \tag{85}$$

Substituting expression (85) into expression (84) gives

$$W = \sum_t \frac{1}{(1+r)^t} \left\{ v_t(p_t) + Q_t(p_t)(p_t - c_t) - \beta_t \left[Q_t(p_t) - (1-d)Q_{t-1}(p_{t-1}) \right] \right\} \; . \tag{86}$$

Maximizing expression (86) with respect to p_t yields the first-order condition

$$\left\{ Q_t'(p_t)(p_t - c_t) - \beta_t Q_t'(p_t) \right\} + \frac{1}{1+r} \left\{ (1-d)\beta_{t+1} Q_t'(p_t) \right\} = 0 \; ,$$

which simplifies to

$$p_t = C_t + c_t$$

where C_t is defined in expression (83). 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.

[116] 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. See, for instance, expression (7) in Biglaiser and Riordan (2000).

Of course, as we emphasize throughout this chapter, the cost structure of the regulated firm and the rate of technical progress are unlikely to be common knowledge in practice. The information that is available about the firm's cost structure and about the firm's potential for achieving productivity gains usually is incomplete in many respects. The regulated firm often will claim that its potential for cost reduction and the rate of technical progress are modest, while consumer advocates will argue that the firm is capable of achieving pronounced cost reductions and productivity gains. The 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.

The regulator can also adopt one of several measures to better cope with the uncertainty he faces. For instance, the regulator might implement an earnings (or profit) sharing plan instead of a pure price cap regulation plan. Under a typical earnings sharing plan, the regulated firm shares with its customers a portion of the earnings it generates above a specified threshold. Although an earnings sharing requirement can limit the firm's incentive to reduce its operating costs, it can also limit the rent that accrues to the firm when its costs turn out to be unexpectedly low.[117]

Alternatively, the regulator might offer the firm a choice among regulatory plans, e.g., a pure price cap plan and an earnings sharing plan. As the analysis in section 2 suggests, a carefully structured choice among regulatory plans can 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 regulation plan with a high X factor (i.e., a high average rate of decline in prices); and (2) an earnings sharing plan with a lower X factor. 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.[118] 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 that it knows it can generate. The less capable firm is willing to share its (relatively modest) earnings with its customers since doing so allows it to guarantee more modest price reductions.

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 this linkage in more detail, and explores the tradeoffs involved in varying the extent to which prices reflect realized costs. The focus in this section is on the tradeoff between allocative efficiency and providing incentives for the firm to control its costs. The discussion proceeds in the moral hazard setting of section 2.4.

Recall from section 2.4.1 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

[117] Lyon (1996) demonstrates that it is generally optimal to link prices to realized costs to some extent. Such linkage can be effected via earnings sharing. See Sappington and Sibley (1992) for related observations.

[118] See Laffont and Tirole (1986) and Lewis and Sappington (1989d) for formal analyses of this issue, and pp.155–165 of Sappington and Weisman (1996a) for further discussion.

entire social gains that its unobserved activities secure. The reason is that 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. And incentive issues are resolved fully when the firm is the residual claimant for the surplus it generates. 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 socially costly.

For simplicity, consider the moral hazard setting where there are no fixed costs of production, and constant unit costs 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, $\hat{\phi}$, is given by expression (52) above. Suppose further that the demand function is iso-elastic, with constant elasticity equal to η.[119] And suppose that transfer payments are prohibitively costly, so the regulator can only dictate the unit price that the firm will be allowed to charge given its realized costs.[120] Therefore, the relationship between consumer surplus and the firm's utility in state $i, V_i(U_i)$, is as specified in expressions (57) and (58).

In this setting, 3.2.3 prices are required to perform two tasks. First, they must provide the firm with incentives to reduce costs, which requires that profits be higher when 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.

The full-information pair of prices, 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, from expressions (58) and (59), given by

$$\frac{p_L - c_L}{p_L} = \frac{p_H - c_H}{p_H} = \left[\frac{\lambda}{1+\lambda}\right]\frac{1}{\eta}\,. \tag{87}$$

Thus, the Lerner index $(p_i - c_i)/p_i$ is equal for the two cost realizations, in accordance with standard Ramsey principles. At this full-information outcome, prices vary directly 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 reduce its costs.

Turning to the second-best problem, where the regulator cannot directly control the firm's

[119] If demand is inelastic so $\eta \leq 1$, suppose that demand goes to zero when price reaches some high "choke price" $\hat{p}$, 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.

[120] Implicitly, we rule out both transfer payments from taxpayers to the firm and two-part tariffs (i.e., transfer payments from consumers to the firm). 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.

cost-reducing effort, expression (60) in the present setting becomes

$$\frac{p_L - c_L}{p_L} = \left[\frac{\hat{\lambda} + \Delta^V \hat{\phi}'}{1 + \hat{\lambda} + \Delta^V \hat{\phi}'} \right] \frac{1}{\eta} \, ; \; \frac{p_H - c_H}{p_H} = \left[\frac{\hat{\lambda} - \Delta^V \hat{\phi}'}{1 + \hat{\lambda} - \Delta^V \hat{\phi}'} \right] \frac{1}{\eta} \, . \tag{88}$$

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, $\hat{\phi}' = 0$ and expressions (88) reduce to the standard full-information Ramsey formula (87).) Thus a form of 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 can be accomplished by setting a price that does not vary with realized costs (so $p_L = p_H$).[121] In this case, it is optimal to implement pure price cap regulation, and the full-information outcome is achieved again.[122]

In less extreme cases, departures from the full-information policy are optimal. Expressions (88) imply that

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

and so the Lerner index is higher in the low-cost state than the high-cost state, in order to provide incentives for cost reduction. In particular, the optimal price does not rise as rapidly as realized costs rise. Indeed, it might even be optimal to set price below cost in the high-cost state to encourage the firm to obtain low costs.

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 serve to motivate the firm to achieve low costs.[123]

3.4 Regulatory Discretion

The final key element of the design of regulatory policy that will be considered is here is the degree of policy discretion to afford the regulator. When the regulator has extensive,

[121] The solution involves $\hat{\lambda} = 0$.

[122] This is essentially an instance of the analysis of optimal regulation with a risk-neutral firm when transfers are used, discussed in section 2.4.1. When demand is perfectly inelastic, there is no difference between the use of prices and lump-sum transfers, and the prohibition on the regulator's use of transfers is not restrictive.

[123] This analysis is closely related to that for the risk-averse firm when transfers are possible, as discussed in section 2.4.2. In both cases, there is a concave relationship between consumer surplus and the firm's utility, and it is this feature that makes the optimal regulatory policy less high powered than the full-information policy.

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, there is always the risk that the regulator might act opportunistically. 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 way. These two dangers are discussed in turn.

3.4.1 Policy credibility

Section 2.3.3 explained how a regulator's inability to commit to future policy can harm the regulatory process. The key problem in section 2.3.3 was that the regulator could not refrain from using information revealed early in the process to maximize future welfare. Another fundamental problem arises in the presence of sunk investments.[124] Once the firm has made costly and irreversible investments, the regulator with limited commitment powers may choose not to compensate the firm for those investments, in an attempt to 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.[125] Anticipating expropriation of some form, the firm will typically undertake too little surplus-enhancing investment.[126]

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.[127] Although this kind of "cost-plus" regulation can provide limited incentives for cost reduction in a static

[124]See Williamson (1975) for a pioneering treatment of the problem, and see chapter 2 in Newbery (2000) for a detailed discussed of the problem of regulatory commitment. Tirole (1986b) considers both the information and investment aspects of the policy credibility problem.

[125]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 where 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.

[126]Another possible response to the threat of expropriation might be for the firm to distort its capital structure. 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.

[127]See Greenwald (1984) for this analysis. See Levy and Spiller (1994) and Sidak and Spulber (1997) for an examination of the legal framework governing a regulator's ability to expropriate a firm's sunk investments.

setting (recall section 3.3), it can provide relatively strong incentives for investment in a dynamic setting where the regulator has weak commitment powers. Nevertheless, 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". However, there are two problems with this modified policy. First, an investment might ultimately prove to be unnecessary even though it was originally desirable. Second, to the extent that the regulator has discretion regarding the particular sunk investments that are included in the asset base, the problem of limited regulatory commitment resurfaces.[128]

Although limited regulatory commitment can discourage investment, it need not always do so.[129] 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.[130] Thus, the dynamic setting, which is the source of the credibility problem, can be used to mitigate the credibility problem. To illustrate, in a model where investments last forever—which is where the danger of expropriation is especially great—desired investment levels can 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.[131]

Another way to mitigate commitment problems may be to divide the overall regulation of the firm among different regulatory bodies with different objectives. When regulatory failure—either in the form of an inability to commit, or a susceptibility to capture—is not an issue, control of the firm by a single body is typically optimal. If the firm is controlled by multiple bodies, each with different objectives, the (equilibrium) outcome may be sub-

[128] 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.

[129] 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.

[130] Of course, this is just an instance of the general theory of dynamic and repeated games. See chapter 5 of Fudenberg and Tirole (1991) for an overview. Gilbert and Newbery (1994) and chapter 2 in Newbery (2000) 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) and Lewis and Sappington (1991b) assess the merits of alternative regulatory charters.

[131] See Salant and Woroch (1992) for a formal analysis of this issue. 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.

optimal because externalities among regulators may be ignored. In particular, one regulator may determine his own policy without regard for its effect on the objectives of another regulator.[132] However, when regulatory failure is possible, the inefficiency caused by policy externalities may act to mitigate the failure.[133]

Finally, one way to lessen politicians' temptation to expropriate sunk investments is to increase the political cost of so doing. For instance, the government might encourage wide participation in its privatizations by setting low initial share prices. If a large fraction of the population has a meaningful stake in the firm, expropriation of the firm (and its many shareholders) may be politically costly.[134]

Much of the preceding discussion has taken for granted the premise that commitment is a good thing. In general, commitment is desirable if the regulator can be trusted to act benevolently. However, if regulatory capture is possible, regulatory commitment need not be unambiguously beneficial, because the ability to commit may facilitate long-lived undesirable policies. To see why, suppose in each period there is a regulator who might (with some exogenous probability) be susceptible to capture by the firm. In this setting, suppose the government can decide whether to allow the regulator to write long-term contracts with the firm, i.e., whether the regulator can commit to future policy. Endowing the regulator with commitment power involves a trade-off: commitment can enable the regulator to promise credibly note to expropriate the firm's sunk investments (and thereby encourages such investments), but it also allows a corrupt regulator to inflict long-run damage on society. Whether commitment is desirable, therefore, depends (in a complicated way) on the various parameters of the model. For instance, commitment is desirable if the probability of capture is small (as one would expect), but commitment can also be desirable if capture is very likely.[135] A second way in which the regulator might not be fully benevolent is if he is myopic. For instance, he 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 ability to write long-term contracts might be undesirable since it could allow

[132] See Baron (1985) for an example of this analysis.

[133] To illustrate this possibility, consider the possible benefits of private versus public ownership of an enterprise as discussed in Laffont and Tirole (1991c). 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 controlling him: the regulator (who is interested in maximizing future welfare) and the 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 high 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 the regulators' temptation to renegotiate contracts over time.

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

[135] Laffont and Tirole (1993, Chapter 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, that there are two periods and that 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 suggests that short-term contracts are desirable.

costs to be passed on to future periods.[136]

3.4.2 Regulatory capture

In the model of regulatory capture analyzed in section 2.2.2, 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 stake in collusion. That model proposed what might be termed a "complete contracting" response to the capture problem, and the regulator was given an explicit monetary incentive by the "constitution" to behave benevolently. In practice, it may be difficult to write a regulatory constitution with such detailed contingencies. Instead, the constitution may only be able to specify broad regulatory objectives and the instruments that can be employed to pursue the objectives. For instance, one response to the danger of capture might be to forbid regulators from future work within the industries they oversee.[137] This section considers this "incomplete contracting" approach to the capture problem.[138]

When explicit monetary incentives cannot be provided to the regulator, it may be desirable to limit his discretion. By precluding the regulator from undertaking actions that are likely to benefit the firm but unlikely to benefit society, the potential losses from capture are lessened, and the incentives of the firm to expend wasteful resources on influencing the regulator are reduced. Of course, restricting the regulator's freedom to act may preclude certain welfare-enhancing actions. Consequently, the regulator should be given full authority to act if he is fully trustworthy.

The theory of optimal regulation summarized in section 2 diverges from common regulatory practice by assuming that the regulator can implement lump-sum transfers to and from the firm. In practice, regulators often rely solely on prices to pursue their objectives. Since the use of transfer payments would typically improve the regulatory process, their absence likely reflects some form of regulatory failure. The failure could entail regulatory capture.[139] To see why, suppose the regulated firm's fixed costs initially are 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.[140] 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

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

[137] 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). See also Salant (1995) for an analysis of how non-contractible investment could be encouraged when the regulator may later be employed by the firm.

[138] See section 15.1 of Laffont and Tirole (1993b) for an expanded discussion of this issue.

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

[140] There is therefore a restriction to linear pricing in the no-transfer case. It is seems likely, though, that the argument can be modified to allow for two-part tariffs.

may be acutely aware of any report of high costs by the firm/regulator, since high costs translate into higher (average cost) prices. If consumers are somehow 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 reduced capture could outweigh the allocative inefficiencies introduced by the use of average-cost pricing.

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.

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. 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.[141]

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.[142] 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.[143]

However, 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 of increasing quality, but the price cap constraint prevents the firm from recovering the

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

[142] 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.

[143] 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.

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.[144]

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.[145] 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.[146] 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 unregulated services arises because the cost allocation procedure effectively taxes the firm's output of unregulated services, which reduces their supply.[147]

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

[144] See Laffont and Tirole (2000, p.88). Spence (1975) and Besanko *et al.* (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.

[145] See Berg and Lynch (1992) and Lynch, Buzas, and Berg (1994).

[146] 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.

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

markets, which explains the excessive expansion of these operations.[148]

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.[149]

Although operations in unregulated markets can harm consumers of regulated services by admitting cost shifting and encouraging inefficient production technologies, diversification into unregulated markets can also 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.[150] 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.[151]

A regulator can also 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.[152]

A regulator can also 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 (1982)'s model in which the regulated firm 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.[153]

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

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

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

[151] See Palmer (1991).

[152] See Braeutigam (1993).

[153] 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.

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 be better able to 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 disposal. The regulator with fewer instruments than objectives typically will be unable to achieve 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 is often 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. We focus on two of these benefits: the *rent-reducing effect* and the *sampling effect*. 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 they may command less rent from their private information. This is the rent-reducing effect of competition. The sampling effect 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 effects 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 effects of competition and others 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. Initial franchise auctions and franchise renewals are both analyzed in section 4.2. 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 a regulated industry. Section 4.4 analyzes the number of firms that a regulator should authorize to produce a single regulated product. Section 4.5 extends this analysis to settings where there are multiple regulated products, and the regulator can determine which firm supplies which products. We will see that there is often a rent-reducing effect of integrating product (i.e., choosing a single firm to supply all products), 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 may render operation by two or more firms 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 give rise to a powerful rent-reducing effect. The effect 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 effect 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 jurisdiction.

4.1.1 Yardstick Performance Setting

To illustrate the benefits of yardstick competition, consider first the following simple "yardstick performance setting".[154] Suppose there are n identical and independent markets, each served by a 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 $i = 1, ..., n$. The regulator specifies the price p_i that firm i must set and the lump-sum transfer 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, taking as given the actions of the other firms. Thus, collusion does not occur in this yardstick performance setting.

Proposition 13 reveals how, despite his limited knowledge, the regulator can exploit the symmetry of the environments to achieve the full-information outcome, i.e., the outcome he would implement if he were perfectly informed about the environment. 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., $Q(c_i) + F'(c_i) = 0$).

Proposition 13 *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,*

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

Since each firm's compensation is independent of its own actions under the reward structure described in Proposition 13, each firm acts to minimize its own total 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 13 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.[155] In principle, corresponding results could be achieved if the producers faced different operating environments. In this

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

[155] 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 VF mechanism in section 3.2.1. There, the repeated observation

case, though, the regulator would require detailed knowledge of the differences in the environments in order to ensure the full-information outcome.[156] 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.[157]

A crucial simplifying feature of the yardstick performance setting is that the firms face no uncertainty. If uncertainly is introduced into the production functions, then the full-information outcome typically is not possible when firms are risk averse. This is because, as discussed in section 3.3, the regulator must consider the firms' aversion to risk when determining the optimal power of the incentive scheme. The policy proposed in Proposition 13 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 base each firm's reward on the performance of other firms, thereby incorporating yardstick competition to some degree.[158]

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 the costs of other firms. Then a reward structure like the one described in Proposition 13 will not induce the full-information outcome because it does not reward each firm adequately for the beneficial impacts of its expenditures on the cost 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.[159]

4.1.2 Yardstick Reporting Setting

Yardstick competition can also admit a powerful rent-reducing effect 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 examined in section 2.2.1.[160] There

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.

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

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

[158] 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 section 3.4 of Armstrong, Cowan, and Vickers (1994) for a simplified analysis in which regulatory policy is restricted to linear schemes.

[159] Dalen (1998) and Sobel (1999) provide proofs of this observation. Meyer and Vickers (1997) provide related insights in their analysis of implicit rather than explicit relative performance comparisons.

[160] This discussion is based on the analysis in Demski and Sappington (1984) and Cremér and McLean (1985).

are two firms, A and B, which 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 marginal cost $c^B \in \{c_L^B, c_H^B\}$ 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 costs truthfully, and consider the optimal policy towards firm A. Let ϕ_i^A denote the probability that firm B has a low-cost realization c_L^B when firm A's marginal cost is c_i^A, for $i = L, H$. To capture the fact that the two firms operate in correlated environments, assume that $\phi_L^A > \phi_H^A$. Just as in section 2.2.1, without any bounds on the penalties that can be imposed on the risk-neutral firm, the regulator can ensure marginal-cost pricing for firm A without ceding the firm any rent. 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_{ij}^A be the lump-sum transfer payment to firm A when it claims its cost is c_i^A and when firm B's cost is c_j^B. If firm A claims to have a high cost, it is permitted to charge the (high) unit price, $p_H^A = c_H^A$. 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_H^A. Formally:

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

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 high price (c_H^A). Formally:

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

The transfers T_{HL}^A and T_{HH}^A can always be set to satisfy this pair of expressions except in the case where the costs of the two firms are independently distributed ($\phi_L^A = \phi_H^A$). Consequently, provided that firm B reports its cost truthfully, the full-information outcome can be implemented for firm A. Firm B's cost report serves precisely the same role that the "audit" did in section 2.2.1.

Notice further that 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 relatively low costs knows that other firms are also likely to have relatively low costs. Consequently, cost exaggeration poses considerable risk of a severe penalty.

When the firms' costs are not highly correlated, substantial penalties are generally required to eliminate a firm's unilateral incentive for cost exaggeration. Just as in section 2.2.1, this can be problematic if firms are risk averse or if feasible payoffs to firms are bounded.[161]

[161] 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.

Another potential complication with a yardstick reporting policy of this type is that it might encourage "collusion" between the firms. Although there is an equilibrium where the two firms truthfully report their private cost information, other equilibria may 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.[162]

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.[163] When multiple potential suppliers are present, both the sampling effect and the rent-reducing effect of competition can arise.

A static model

To illustrate how a regulator might employ franchise bidding to discipline a monopoly supplier, consider the following single-period setting based on the Baron-Myerson model described in section 2.1.1. Suppose there are $N \geq 1$ firms that are qualified to serve as a monopoly provider in a particular market.[164] Each firm has either low marginal cost (c_L) or high marginal cost (c_H). As usual, the probability that a given firm has a low cost realization is ϕ, and this outcome is realized independently across the N firms. The firm 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.[165]

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.[166] When a firm that truthfully announces cost c_i is selected to produce, it will receive rent $R_i = Q(p_i)(p_i - c_i) - F + T_i$. However, a firm that announces cost c_i will only be selected to produce with some probability, ρ_i. In the equilibrium where all firms announce their costs

[162] 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 complications that arise when regulated firms are able to coordinate their actions explicitly.

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

[164] 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.

[165] The possibility of simultaneous production by multiple producers is considered below in section 4.3, as is the possibility of an endogenous number of active producers.

[166] 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.

truthfully (which can be considered without loss of generality if there is no collusion between firms), a high-cost firm 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 equilibrium probability that a given high-cost firm will win the auction. Similarly, if a firm has low costs, it will win the contest with the (higher) probability[167]

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

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 (2), 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 $\rho_L R_L \geq \rho_H[R_H + \Delta^c Q(p_H)]$, or

$$R_L \geq \frac{\rho_H}{\rho_L}\left[R_H + \Delta^c Q(p_H)\right] \ . \tag{89}$$

Comparing expression (89) with expression (2), the corresponding constraint when there is only one potential supplier, it is apparent that competition relaxes the relevant incentive compatibility constraint.[168] This is the rent-reducing effect of competition.

As in expression (13), 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 $\alpha \leq 1$ 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

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

Comparing this expression with expression (14), the corresponding expression when there is only one potential producer, reveals another beneficial effect of competition: the probability that the monopoly producer has low costs increases. This is the sampling effect of competition.

Standard arguments show that $R_H = 0$ and $p_L = c_L$ under the optimal policy. Also, the incentive constraint (89) will bind, and so p_H will be chosen to maximize

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

The maximization provides:

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

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

[168] As usual, the only binding incentive compatibility constraint is the "downward" constraint that ensures the low-cost firm will not exaggerate its costs.

which does not depend on N, and is exactly the optimal price in the absence of competition for the market, as given in expression (3).

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.[169] This invariance holds because two conflicting effects offset each other exactly. 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 chances (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 as N increases, the likelihood that a low-cost firm will be awarded the franchise increases. Therefore, it becomes more important to reduce the rent of the low-cost firm by raising p_H above c_H. It turns out that these two effects offset each other exactly in this setting with risk-neutral firms with independently distributed costs.

Expression (89) 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 the high-cost price p_H is independent of the number of bidders, this rent decreases with the number of bidders.[170] Furthermore, since the overall probability that a low-cost firm wins is $[1-(1-\phi)^N]$, the aggregate expected rent of all bidders is:

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

This expected industry rent is decreasing in N. These key features of the optimal regulatory policy in this setting are summarized in Proposition 14.[171]

Proposition 14 *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 costs.

(ii) A high-cost firm makes zero rent.

(iii) The rent enjoyed by a low-cost firm that wins the contest decreases with the number of bidders.

(iv) The total expected rent of the industry decreases with the number of bidders.

(iv) 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 given in (3).

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,

[169] 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'.

[170] 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.

[171] 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.

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.[172,173] 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.

Dynamic considerations

Although franchise bidding admits the beneficial rent-reducing and sampling effects of competition, it is not without its potential drawbacks. These drawbacks include the following three.[174] 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 monopolist 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.[175]

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 supplier can benefit from its investments. To alleviate the tension introduced by these two countervailing 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 supplier 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 supplier will be selected to operate the franchise in the future can increase the supplier's expected return from such transferable, sunk investments. Consequently, such a bias can enhance incentives for the

[172] For instance, see the discussion in section VII of McAfee and McMillan (1987a).

[173] 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).

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

[175] 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.

incumbent supplier to undertake these valuable investments.[176] 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 lower operating costs than its rivals in subsequent auctions, it can be optimal to bias the subsequent auctions *against* the incumbent.[177]

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 supplier to an alternative producer. The second source might be a firm that presently serves other markets, or it might be a potential producer that does not presently operate elsewhere. Second sourcing can increase welfare in two important ways. It can do so directly by shifting production from the incumbent supplier to the second source when the latter has lower operating costs than the former (the sampling effect). It can also do so by reducing the rent that the producer secures from its privileged knowledge of its operating environment. This rent-reducing effect can arise from two distinct sources. First, as reflected in expression (89) above, the incumbent producer will be less inclined to exaggerate its operating costs when the probability that it is permitted to operate declines as its reported costs increase.[178] 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.[179]

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 policy described in section 2.2.1 (as well as the yardstick reporting policies of section 4.1.2). In particular, an incumbent that reports high costs 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 implicitly corroborates the incumbent's report by reporting high costs 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 cost 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 supplier, it can

[176] 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.

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

[178] 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 producer with a second source, the quantity distortions that are implemented to limit information rents may be reduced. 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.

[179] 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).

be optimal to use the second source even when it is known to have higher costs than the incumbent.[180]

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 firm earns when its operating costs are low. When it anticipates little or no rent from realizing low production costs, the incumbent firm 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, the *ex ante* elimination of a second source helps to restore some of the commitment power that is needed to motivate cost-reducing effort.[181]

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.[182] In these settings, the presence of alternative unregulated producers can affect the optimal regulation of the dominant firm, and overall welfare, in a variety of ways. The effect of competition on welfare can be positive or negative. In particular, competition can introduce the beneficial rent-reducing and sampling effects described above. However, unregulated competitors may undermine socially desirable tariffs that have been imposed on the regulated supplier.

To analyze these effects formally, consider the following simple example, which extends the Baron-Myerson model summarized in section 2.1.1. 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 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 that the fringe always offers the product at price c^R. (For simplicity, we abstract from fixed costs of production for the fringe.)

There are four cases of interest. First, suppose $c^R < c_L$. The fringe will increase welfare in this case, because the industry 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 $p = c_L$. In return, the firm is paid a subsidy equal to its fixed costs. 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-

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

[181] 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.

[182] 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.

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 beneficial sampling effect and a beneficial rent-reducing effect in this setting. The sampling effect arises because the fringe supplies the market at lower cost than can the high-cost dominant firm. The rent-reducing effect arises because the dominant has no freedom to exaggerate its costs. Whenever the dominant firm has or claims to have high costs, it is replaced by the fringe as the industry supplier.

Third, suppose $c^R > p_H$, where p_H is given in expression (3). 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 possible policies: (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 costs. Policy (ii) is implemented by providing the dominant firm with only a single alternative to shutdown: set price equal to c_L and receive a subsidy equal to the firm's fixed costs of production.

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. This fact is evident from expressions (26) and (27). Recall that once the rents of the dominant firm are accounted for, expected welfare is the welfare derived from marginal cost pricing with fully-adjusted costs (e.g., $\hat{c}_H = c_H + \left[\frac{\phi}{1-\phi}\right][1-\alpha]\,\Delta^c$). Because the fringe has a lower marginal cost than the adjusted cost of the high-cost dominant firm ($c^R < \hat{c}_H$), expected welfare is higher in this case when the fringe operates in place of the high-cost dominant firm.[183]

Proposition 16 summarizes these observations for this competitive fringe setting (where the fringe is unregulated, but its production costs are known).

Proposition 15 *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, but where the fringe's

[183] This same logic explains why the regulator might favor a less efficient bidder in a franchise auction, as discussed in section 4.2 above.

cost is uncertain, and may be correlated with the dominant firm's cost.[184,185] 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 that the incumbent regulated firm has no private information about its cost of operation. The central friction in this setting arises because (in contrast to the other settings considered above) a deadweight loss is incurred whenever funds are transferred from consumers to the regulated firm. The deadweight loss might reflect the distortions that arise when taxes are imposed to generate the funds required to make transfer payments to the regulated firm.[186] For simplicity, suppose $\$(1+\Lambda)$ is the cost of transferring \$1 to the firm, where $\Lambda > 0$. Further suppose the regulator values consumer surplus and the rent of the regulated firm equally. The firm offers n products at prices $\mathbf{p} = (p_1, ..., p_n)$. At these prices, the firm's profit is $\pi(\mathbf{p})$ and consumer surplus is $v(\mathbf{p})$. In this setting, welfare with prices $\mathbf{p}$ is

$$W(\mathbf{p}) = v(\mathbf{p}) + (1+\Lambda)\pi(\mathbf{p}) \, . \tag{90}$$

In the absence of competition, optimal (Ramsey) prices $\mathbf{p}^*$ will simply be chosen to maximize $W(\cdot)$.

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 may not interfere with the Ramsey prices that maximize expression (90). 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 marginal cost is sufficiently smaller than the corresponding marginal 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 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 simply limit the options available to the regulator without offering offsetting benefits, such as those that arise from the rent-

[184] 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 costs when the regulated firm does. Consequently, the reduced output that the regulated firm will be authorized to produce when it exaggerates costs 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.

[185] 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.

[186] If transfers were costless, the regulator could ensure the ideal full-information outcome simply by setting marginal cost prices and delivering the transfer payment required to ensure the firm's operation. Competition would only be beneficial in such a setting.

reducing or sampling effects of competition.[187,188]

Such undesirable entry also can occur when the regulator has distributional objectives, and favors the welfare of one group of consumers over another.[189] (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 (e.g., long distance and international calls) 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 price above marginal cost that can arise under simple 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 competition can be to undermine the tax base. This perspective suggests an obvious solution to the problem caused by unregulated competition: require consumers to pay the same implicit tax whether they purchase a product from the regulated firm or the competitive fringe. Such a policy, which entails regulation of the fringe, can ensure that entry occurs only when the fringe 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. We discuss this kind of policy in section 5.1.1 below.

In practice, it is often impractical to levy taxes directly on the products supplied by competitors. In some settings, though, access charges can be employed to levy such taxes indirectly. This possibility is considered in sections 5.1.2 and 5.1.3 below.

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

---

[187] 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.

[188] 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 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.

[189] See chapter 6 of Laffont and Tirole (2000) and Riordan (2002) for discussions of this issue, and for further references.

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 optimal 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.[190]

When choosing the number of firms to operate in an industry, a fundamental tradeoff often arises. Although additional suppliers can introduce favorable competitive effects (such as increased product variety and quality, and the rent-reducing and sampling effects of competition discussed above), industry production costs can increase when production is dispersed among multiple suppliers.

To examine how the tradeoff is optimally resolved in a regulated setting, consider the following simple variant of the Baron-Myerson model of section 2.1.1.[191] If the incumbent firm faces no competition, the optimal regulatory policy is as described in Proposition 1. Recall that this policy delivers rent to the firm when it has low costs. The policy also implements a price in excess of marginal cost when the firm has high costs.

Now suppose that the regulator can, if he so chooses, license a rival firm to operate in the market. Further suppose that the rival's marginal cost is always the same as the incumbent's marginal cost. If the rival enters the market, the two firms engage in Bertrand price competition. 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. Of course, anticipating the intense competition that will ensue, the rival 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 the welfare losses that arise from asymmetric information about the incumbent firm's operating costs. The regulator will purchase this instrument only if the benefits it provides outweigh its cost, which is the rival's fixed operating cost (F).[192] If F is sufficiently small, the regulator will induce the rival to operate. Therefore, since the regulator would always authorize only a single supplier in the absence of asymmetric information about operating costs, such asymmetric information renders the regulator more likely to introduce competition into the regulated industry.

Now 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 princi-

[190] In addition, 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 below.

[191] The following discussion is based on section 4.1.1 of Armstrong, Cowan, and Vickers (1994).

[192] See Auriol and Laffont (1992) and Riordan (1996) for formal proofs and more detailed explanations of this and related observations in models where the firms' costs are not perfectly correlated.

ple, entry restrictions could increase welfare.[193] In practice, of course, it is a non-trivial 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.[194]

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 consumers), a regulator may intentionally restrict the number of licenses that he issues, thereby creating a relatively concentrated industry structure.[195]

Entry policy also can affect the speed with which consumers are served. Consider, for example, a setting where firms must incur fixed, sunk costs in order to operate, and where firms have different marginal costs of production. If a regulator were simply to authorize a single, randomly-selected firm to operate, redundant fixed operating costs could 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.[196]

To this point, the discussion 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 policy maker relies on advice from a (better informed) reg-

[193] 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 produce 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 a phenomenon that is similar to the sampling effect of competition).

[194] McMillan (1994), McAfee and McMillan (1996), Cramton (1997), Milgrom (1998), and Salant (2000) 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 licences to *two* firms, who subsequently compete to innovate.

[195] 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.

[196] 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.

ulator 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.[197]

4.5 Integrated Versus Component Production

In multiproduct industries, regulators often face the additional 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 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 informational economies of scope, first consider the following simple setting with independent products.

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.[198] Suppose further that each product has 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 under integrated production. To see why, suppose the integrated firm is regulated according to the regime suggested in Loeb and Magat (1979), so that the firm is free to set the price it charges for each of its product, and the firm keeps the entire consumer surplus that its price structure generates. For the reasons identified in section 2.1.1, the firm will set prices equal to marginal costs under this regulatory policy. 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 because 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 equal to its expected rent, thereby approximating

[197] See Laffont and Tirole (1993a) for a formal analysis of this effect. Thus, the possibility of capture, which might be expected to reduce the likelihood of entry, acts to *increase* the likelihood of entry once the political principal has responded appropriately to the threat. A similar observation was made in section 2.2.2 above.

[198] The following discussion, found in Dana (1993), also applies naturally to the subsequent discussion about complementary products.

the full-information outcome.

In this simple, but extreme, 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 2.2.3.[199] 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. Then 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 producer of each component. It is always possible for the regulator to choose this regime under integrated production. However, part (iii) of Proposition 5 shows 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 shows that the full-information outcome is possible with yardstick competition, and so component production is always optimal, provided the two producers do not collude. In contrast, the full-information outcome will not be attainable if the firms must receive non-negative rent for all cost reports (de 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.[200,201] (When the correlation between costs is high, part (ii) of Proposition 5 shows 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.)

[199] The following discussion is based on Dana (1993).

[200] 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.

[201] 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 cost are negatively correlated.

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 areas, 1 and 2, and the regulator must decide whether to auction the two areas in separate auctions or to "bundle" the areas together in a single franchise auction. Suppose there are two potential operators, A and B, each of which can operate in one or both areas. Suppose the cost of providing the specified service in area k is c_i^k for firm i, where $k = 1, 2$ and $i = A, B$. Further, suppose there are no economies (or dis-economies) or scope in joint supply, so that firm i's cost of supplying both areas is $c_i^1 + c_i^2$. Suppose the regulator wishes to ensure production in each area, and so imposes no reserve price in the auction(s).

If the regulator awards the franchise for the two areas in two separate second-price auctions, he will have to pay the winner(s)

$$\max\{c_A^1, c_B^1\} + \max\{c_A^2, c_B^2\} \ . \tag{91}$$

If the regulator awards the two areas in 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 (91). Therefore, the regulator will pay less when he bundles the two franchise areas in a single auction than when he conducts two separate auctions (with potentially two different winners). This conclusion reflects the rent-reducing effect of integrated production.[202]

Complementary products

Now, suppose there is a single final product that is produced by combining two essential inputs.[203] 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.[204]

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,

[202] This discussion is based on Palfrey (1983), who shows that with more than two bidders, the ranking between integrated and component production may be reversed. Notice that when the two areas 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 analyses of the optimal design of auctions with multiple products, see, for example, Armstrong (2000) and Avery and Hendershott (2000).

[203] The following discussion is based on the analysis in Baron and Besanko (1992) and Gilbert and Riordan (1995).

[204] 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.

c_L or c_H, where $c_L < c_H$. The probability of obtaining a low-cost outcome is ϕ, and the costs of producing the two inputs are independently distributed. Suppose it is optimal to supply one unit of the product except when *both* of its inputs have a high cost.[205]

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 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 (40), 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 $R_{LL} \geq \Delta^c \equiv c_H - c_L$. 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 that if a firm reports that it has cost $c = c_i$ it receives the expected lump-sum payment $\bar{T}_i$. 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 $\bar{R}_L = \bar{T}_L - c_L$. If a firm reports that it has high costs, then production takes place only with probability ϕ (i.e., when the other firm has low costs), and so the firm's expected rent is $\bar{R}_H = \bar{T}_H - \phi c_H$. The regulator will ensure that a firm receives no rent when it has high costs, so $\bar{R}_H = 0$. Furthermore, the minimum rent that ensures truthful revelation of low costs is $\bar{R}_L = \phi \Delta^c$. (When it reports high costs, a firm risks being shut down and earning no rent with equilibrium probability $1 - \phi$. Consequently, the equilibrium expected rent of a firm with low costs is only $\phi \Delta^c$.) Therefore, the regulator must deliver an expected rent of $\phi^2 \Delta^c$ to each firm under component production, yielding a total expected 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.[206]

The regulator's preference for integrated production in this setting arises because integration serves to limit the firm's incentive to exaggerate its costs. 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 costs are too high. Termination reduces the profit that can be generated on both inputs. Under component production, a firm that exaggerates its operating costs risks only the profit that it might secure from

[205] 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.

[206] 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.

producing a single input. Each supplier ignores the potential loss in profit that 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.

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.[207] As such, the result for complementary products is perhaps less surprising than the corresponding result for independent products.

Substitute products

Finally, suppose there are two products, 1 and 2, that consumers view as being perfect substitutes. The cost of producing product k is c^k. This cost can again take one of two values, c_L or c_H. Suppose the probability of a low cost realization is ϕ and the production costs for the two products are independently distributed. The regulator wishes ensure supply of at least one product, and is considering whether to mandate integrated production (where a single firm can supply either of the two products) or component production.

Under integrated production, given that the regulator wishes to ensure the certain supply of one product, he must pay the firm a transfer equal to c_H. In this case the firm makes a rent of Δ^c unless both of its products have high cost. Consequently, the integrated firm's expected rent is

$$R_{INT} = \left(\phi^2 + 2\phi(1-\phi)\right)\Delta^c\ .$$

Under component production, the regulator can ensure the supply of one product by, for instance, auctioning the right to supply to the highest bidder. In this case, there is no rent whenever the two firms have the same costs. When one firm has high cost and the other has low cost, the low-cost firm receives a rent equal to Δ^c. Therefore, the total expected rent under component production is no more than[208]

$$R_{COMP} = 2\phi(1-\phi)\Delta^c\ .$$

Total rent clearly is lower under component production than under integrated production. This is the case because no rent is paid under component production when both products have a low cost realization. Thus, the rent-reducing effect of competition leads to a strict preference for component production—i.e., for competition—over integrated production when products are substitutes.

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

[207] See Cournot (1927).

[208] In fact, the regulator can pay less rent than this to a low-cost firm under component production. For instance, if when both firms report high costs, production is randomly assigned to one firm, then the low-cost firm faces the possibility of not producing when it exaggerates its cost. Consequently, when one firm has high cost and the other has low cost, the low-cost firm receives rent $\frac{1}{2}\Delta^c$. This modification would amplify the preference for component production.

be preferred to integrated production when the costs of producing inputs are highly correlated. This is the case because when costs are highly correlated, the yardstick competition that component production admits can limit rents effectively. Second, integrated production will tend to be preferred to component production when the components are better viewed as complements than as substitutes. In this case, integrated production can avoid what might be viewed as a double marginalization of rents that arises under component production.[209,210]

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.[211]

The regulator's task is more difficult when the 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 are due to the limited service quality that 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 competitive bidding procedures.[212]

---

[209] See Severinov (2003) for a more detailed analysis of the effects of input substitutability or complementarity on the relative merits of component and integrated production. Cost information is assumed to be uncorrelated across the two activities, and so there is no scope for yardstick effects to work. The paper also discusses the alternative industry configuration of "delegation", where the regulator deals with one firm who sub-contracts with the second firm.

[210] Iossa (1999) analyzes a model where the information asymmetries concern consumer demands rather than costs 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.

[211] See Che (1993), Cripps and Ireland (1994) and Branco (1997) for details.

[212] Manelli and Vincent (1995) derive this conclusion in a setting where potential suppliers are privately informed about the exogenous quality of their product. The authors' conclusion that it is optimal to assign the same probability of operation to all potential suppliers is related to the conclusion in section 2.1.3 regarding the optimality of pooling. In Manelli and Vincent's model, incentive compatibility considerations imply that a firm with a low quality product, and thus low operating costs, must be selected to operate at least as

The fact that quality is unverifiable 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 service quality and reasonably low prices arise in equilibrium.[213]

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 (via the sampling effect) and reduce the rents of industry operators (via the rent-reducing effect of competition). Second, competition can complicate the design of regulatory policy considerably. For example, 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. 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.

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.[214] 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 5 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,

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).

[213] 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. When the regulator's information is more limited, though, he typically will prefer to allow competitive forces to determine the customers that each firm serves. See Wolinksy (1997) for an analysis of this issue.

[214] For an account of the theory of vertical relationships in an unregulated context, see Rey and Tirole (2003).

whether the monopolist's retail tariff is regulated, 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.

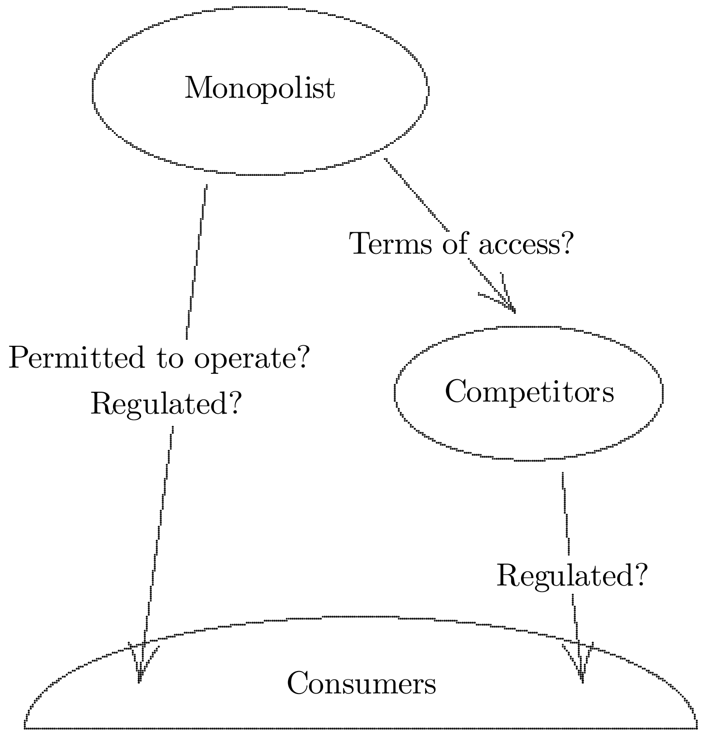

Figure 5: Vertical Relationships

5.1 Access Pricing

Before analyzing (in section 5.1.2) the optimal access policy when the monopoly supplier of the input (access) is vertically integrated, consider the simpler case where the input supplier does not operate downstream.[215,216] 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, in this setting, the markup of the retail price over the total cost of providing the end-to-end service will be close to zero. (The competitive fringe model on which we focus in the rest of this section is, by construction, perfectly competitive, and so setting the access charge equal to the cost of providing access is optimal when the monopolist does not operate in the retail market.) If the downstream market is not perfectly competitive, then it may be optimal (if feasible) to price access

[215] See, for instance, section 5.2.1 in Armstrong, Cowan, and Vickers (1994) and section 2.2.5 in Laffont and Tirole (2000).

[216] See Armstrong (2002) for a more detailed account of the theory of access pricing, from which section 5.1 is taken. Armstrong (2002) also discusses the issue of "two-way" access pricing, where several firms need to obtain inputs from each other. In this section, we abstract from the possibility that the monopolist may try to disadvantage downstream rivals using various non-price instruments. See section 5.2 for a discussion of this point.

below cost in order to induce lower downstream prices, which exceed marginal costs due to the imperfect competition.

5.1.1 The Effect of Distorted Retail Tariffs

The retail prices charged by regulated firms often depart significantly from underlying marginal costs. (As shown below, the access pricing problem would be trivial if this were not the case.) As mentioned in section 4.3, there are two primary reasons why prices might differ from costs. First, in the presence of fixed and common costs, marginal-cost pricing will not allow the incumbent to earn non-negative profit. Consequently, Ramsey prices may be set to implement *optimal* departures of prices from costs. (See section 5.1.3 below.) Second, a regulated firm's retail prices may be set to achieve other goals, such as income redistribution or universal service. In particular, profits from one market may be employed to subsidize losses in other markets. This section discusses the impact of this latter kind of distortion on entry and welfare.

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 prices there are reduced to the level of the competitors' operating costs, and the fringe makes no profit.[217]

Suppose that the fringe and the regulated firm offer differentiated products to final consumers. Let P and p be the regulated firm's price and the fringe's price for their respective retail services. (Throughout this section, variables that pertain to the dominant firm will be indicated by upper-case letters. Variables that pertain to the fringe will be denoted by lower-case letters.) Let $V(P,p)$ be total consumer surplus when prices P and p are offered. The surplus function satisfies the envelope conditions $V_P(P,p) = -X(P,p)$ and $V_p(P,p) = -x(P,p)$, where X and x are, respectively, the demand functions for the services of the regulated firm and the fringe. (Subscripts denote partial derivatives.) Assume that the two services are substitutes, so $X_p = x_P \geq 0$. The incumbent has constant marginal cost C and the fringe has marginal (and average) cost c. In order to achieve the optimal output from the fringe, suppose the regulator levies a per unit output tax t on the fringe's service. Then competition implies that the fringe's equilibrium price is $p = c + t$. Suppose the regulated firm's price is fixed exogenously at P. Suppose further that the regulator aims to maximize total unweighted surplus (including tax revenue).[218] This total surplus is

$$W = \underbrace{V(P,c+t)}_{\text{consumer surplus}} + \underbrace{tx(P,c+t)}_{\text{tax revenue}} + \underbrace{(P-C)X(P,c+t)}_{\text{regulated firm's profits}} . \tag{92}$$

Maximizing W with respect to t implies that the optimal fringe price and output tax are

[217] If entrants did have market power then access charges should be chosen with the additional aim of controlling the retail prices of entrants. This would typically lead to access charges being set lower than otherwise, following the same procedure as the familiar Pigouvian output subsidy to counteract market power. See section 3.3.1 of Laffont and Tirole (2000) for a discussion of this issue.

[218] Since there is no asymmetric information in this analysis, there is no reason to leave the monopolist with rent, and hence maximization of total surplus is an appropriate objective.

given by

$$p = c + \sigma_d(P - C) \; ; \; t = \sigma_d(P - C) \; , \tag{93}$$

where

$$\sigma_d = \frac{X_p}{-x_p} \geq 0 \tag{94}$$

is a measure of the substitutability of the two retail services. In particular, σ_d measures how much the demand for the regulated firm's service decreases when the fringe supplies one additional unit of its service. Equation (93) implies that when sales are profitable for the regulated firm, i.e., when $P > C$, it is optimal to raise the fringe's price above cost as well, i.e., to set $t > 0$. This is because profits are socially valuable, and when $P > C$ 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 (where $t = 0$) 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.

In expression (93), the tax t is set equal to the profit that the regulated firm foregoes when fringe supply increases by a unit. This lost profit is the product of two terms: the marginal profit $(P - C)$ per unit of its final product sales, and σ_d, which is the reduction in its final sales caused by increasing fringe output by one unit. If the services are not close substitutes, so that σ_d is close to zero, then this optimal tax should also be close to zero, and a *laissez-faire* policy towards rivals is nearly optimal. This is because policy towards the fringe's service has little impact on the welfare generated in the regulated firm's market, and therefore there is little benefit from imposing a price in the fringe's market that differs from cost.

The rule (93) is an instance of the theory of the second best. This theory states that if one service is not offered at the first-best marginal cost price ($P \neq C$), then the optimal price in a related market also departs from marginal cost ($p \neq c$). In this sense, the tax in (93) constitutes a second-best output tax. Given the presumed social welfare function (92), it makes little difference whether the proceeds from this tax are passed directly to the regulated firm, to the government, or into an industry fund. However, if the firm has historically been using the proceeds from a profitable activity to finance loss-making activities, then if the fringe pays the tax to the incumbent, the incumbent will not face funding problems as a result of the fringe's presence.[219]

5.1.2 Access Pricing With Exogenous Retail Prices for Incumbent

Now return to our primary focus on vertically-related markets, where the fringe requires access to inputs supplied monopolistically by the regulated firm. In this section, we focus on the problem of how best to determine access charges for a given choice of the regulated firm's retail tariff (which is assumed to be the result of an exogenous regulatory process).

[219] However, perhaps a more transparent mechanism would be for a "universal service" fund to be used to finance loss-making services. See section 2.1 of Armstrong (2002) for further details. More generally, see Braeutigam (1979, 1984) and chapter 5 of Laffont and Tirole (1993b) for discussions of Ramsey pricing in the presence of competition, including cases where rivals are regulated.

It would generally be preferable for the regulator to set the firm's retail prices and access charges simultaneously, since doing so would permit direct consideration of tradeoffs between consumer welfare and productive efficiency. (See section 5.1.3 for this Ramsey analysis.) However, it is instructive to analyze this setting with exogenous (and perhaps inefficient) retail tariffs, because a regulated firm's retail tariffs often are not set according to strict Ramsey principles, as various political, historical, or social considerations often influence retail tariffs.

Suppose that the regulated firm supplies its retail service at constant unit cost C_1, and supplies its access service to the fringe at constant unit cost C_2. As in the previous section, P is the (exogenous) price for the firm's retail service. Let a denote the per-unit charge paid by the fringe for access to the firm's input. Suppose that when it incurs access charge a 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 of 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 $\psi(a) = a + c$, 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, by Shephard's Lemma, the fringe's demand for access per unit of its retail service. Therefore, when it supplies x units of service to consumers, the fringe's total demand for access is $\psi'(a)x$.

The following analysis proceeds in two stages. First, we derive the optimal policy in the case where the regulator has a full range of policy instruments with which to pursue his objectives. Second, we analyze the optimal policy in the setting where the regulator's sole instrument is the access charge.

Regulatory control of fringe output

Suppose, first, that 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 (92), total welfare is

$$W = \underbrace{V(P, t + \psi(a))}_{\text{consumer surplus}} + \underbrace{(P - C_1)X(P, t + \psi(a))}_{\text{monopoly's profits from retail}}$$

$$+ \underbrace{(a - C_2)\psi'(a)x(P, t + \psi(a))}_{\text{monopoly's profits from access}} + \underbrace{tx(P, t + \psi(a))}_{\text{tax revenue}} . \qquad (95)$$

Since $p = t + \psi(a)$, the regulator can be viewed as choosing p and a rather than t and a. In this case, (95) simplifies to

$$W = V(P, p) + (P - C_1)X(P, p) + (p - \{\psi(a) - (a - C_2)\psi'(a)\})\, x(P, p) . \qquad (96)$$

The term in $\{\cdot\}$ brackets in expression (96) is the total cost of producing a unit of the fringe's output when the access charge is a. Since a does not affect any other aspect of welfare in (96), it follows that a should be chosen to minimize this cost $\{\cdot\}$. The relevant first-order condition is $(a - C_2)\psi''(a) = 0$. Therefore, whenever the fringe has some ability to substitute away from the regulated firm's access service, i.e., when $\psi'' \neq 0$, the optimal policy entails marginal-cost pricing of access: $a = C_2$. Also, maximizing (96) with respect to $p = t + \psi(a)$ yields formula (93) for t. In sum, the optimal policy involves

$$a = C_2 \; ; \; t = \sigma_d(P - C_2) \; . \tag{97}$$

Whenever 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 the second-best output tax given in (93). In contrast, if the fringe had access to the regulated firm'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 supply by the fringe 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.

Provided there are enough policy instruments available to pursue all relevant objectives, there is no need to sacrifice productive efficiency even when the regulated firm's retail price differs from its cost. Retail instruments—in the form of an output tax on rivals, for instance—should be used to combat retail-level distortions such as mandated tariffs that are not cost-based. Wholesale instruments should then be used to combat potential productive inefficiencies—in this case the productive inefficiency caused by pricing access other than at cost.

Unregulated fringe output

Now consider the optimal policy when the access charge is the sole instrument available to the regulator. In this case, $t = 0$ in (95), and so welfare under access charge a is

$$W = \underbrace{V(P, \psi(a))}_{\text{consumer surplus}} + \underbrace{(P - C_1)X(P, \psi(a))}_{\text{monopoly's profits from retail}} + \underbrace{(a - C_2)\psi'(a)x(P, \psi(a))}_{\text{monopoly's profits from access}} . \tag{98}$$

Notice that in this setting, the only way the regulator can ensure a high price for the fringe's output (perhaps for the second-best reasons outlined in section 5.1.1) is to set a high charge for access, which will then typically cause some productive inefficiency in fringe supply.

Maximizing expression (98) with respect to a shows that the optimal access charge is

$$a = C_2 + \sigma(P - C_1) \; , \tag{99}$$

where

$$\sigma = \frac{X_p \psi'(a)}{-z_a} \tag{100}$$

and $z(P, a) \equiv \psi'(a)x(P, \psi(a))$ is the fringe's equilibrium demand for access. The parameter σ measures the reduction in demand for M's retail service caused by supplying the marginal unit of access to the fringe.[220] Therefore, expression (99) states that the access charge should be set equal to the cost of access plus the incumbent's foregone profit caused by supplying a unit of access to its rivals. This rule is known as the "efficient component pricing rule" (or ECPR).[221]

[220] For one further unit of access to be demanded by the fringe, a must fall by $1/z_a$, and this induces X to fall by $X_p \psi'/z_a$.

[221] 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) for further discussions of the ECPR.

In the special case where consumer demand for the two retail services are approximately independent (so $X_p \approx 0$), formula (99) states that the access charge should involve no mark-up over the cost of providing access, even if $P \neq C_1$. In other cases, however, the optimal access charge 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 charge, is forced to perform two functions, and the regulator must compromise between productive and allocative efficiency.

This analysis is simplified in the special case where the fringe cannot substitute away from the monopolist's input, so that $\psi(a) = c + a$. In this case, expression (99) becomes

$$a = C_2 + \sigma_d(P - C_1) ,$$

where σ_d is the demand substitution parameter given in (94). This expression states that the optimal access charge is the sum of the cost of providing access and the optimal second-best output tax as given in expression (93). Thus, an alternative way to implement the optimum in this case would be to price access at cost (C_2) and simultaneously levy a second-best tax on the output of rivals, as in expression (97). 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. More generally, however, a strictly higher level of welfare can be achieved if the regulator can use the twin instruments of an output tax and an access charge.

5.1.3 Ramsey Pricing

Having discussed how to set access charges to maximize welfare given the established retail tariff, we now analyze the optimal simultaneous choice of the regulated firm's retail and access prices.[222] As before, the form of the solution will depend on whether rivals can substitute away from the input, and, if they can, on the range of policy instruments available to the regulator.

Regulatory control of fringe output

Suppose first that, in addition to setting the regulated firm's retail price, the regulator can impose a per-unit output tax t on the fringe and set a per-unit charge a for the input. As before, the price of the fringe's service is equal to the perceived marginal cost, so $p = t + \psi(a)$. Suppose also that the proceeds of the output tax are used to cover the regulated firm's fixed costs. As in expression (96) above, the regulator can be considered to choose p and a rather than t and a. Letting $\lambda \geq 0$ be the Lagrange multiplier on the regulated firm's profit constraint, the regulator's problem is to choose P, p and a to maximize

$$W = V(P,p) + (1+\lambda)\left[(P - C_1)X(P,p) + (p - \{\psi(a) - (a - C_2)\psi'(a)\})\, x(P,p)\right] . \quad (101)$$

For given retail prices, P and p, the access charge a does not affect consumer surplus. Consequently, a must again be chosen to minimize the cost of providing the fringe's service, which is the term $\{\cdot\}$ in expression (101). As before, whenever the fringe can substitute away from the input at all (i.e., whenever $\psi'' \neq 0$), the optimal policy is to price access at

[222] See Laffont and Tirole (1994).

cost (so $a = C_2$).[223] The two retail prices, P and p, are then chosen to maximize consumer surplus subject to the regulated firm's profit constraint.

Unregulated fringe output

Next, suppose the regulator has a more limited set of policy instruments. In particular, suppose the output tax t is not available. In this case, $p = \psi(a)$ and the access charge must perform two functions: it must attempt to maintain productive efficiency (as before) and influence the fringe retail price in a desirable way. Following the same logic that underlies expressions (98) and (101), welfare in this setting can be written as

$$W = V(P, \psi(a)) + (1+\lambda)\left[(P - C_1)X(P, \psi(a)) + (a - C_2)\psi'(a)x(P, \psi(a))\right] \; .$$

Letting $\theta = \lambda/(1+\lambda) \geq 0$, the first-order conditions for maximizing this expression with respect to a is

$$a = \underbrace{C_2 + \sigma(P - C_1)}_{\text{ECPR charge}} + \frac{\theta a}{\eta_z} \; , \tag{102}$$

where σ is as given in expression (100), and $\eta_z = -az_a/z > 0$ is the own-price elasticity of the demand for access. Expression (102) states that the optimal access charge is given by the ECPR expression (99), which applies if P were exogenously fixed, 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 charge is called upon to perform too many tasks, a compromise must be made.

In the next section the access charge is forced to perform one further task—to control the dominant firm's retail price.

5.1.4 Unregulated Retail Prices

In this section we discuss how best to price access when the access charge is the regulator's only instrument for controlling the dominant firm, which is now assumed to be free to set its retail price P.[224] For simplicity, suppose there is no output tax on the fringe. As before, if the regulator sets the access charge a, the fringe's price is $p = \psi(a)$. The dominant firm will then set its retail price P to maximize its total profit, which is

$$\Pi = (P - C_1)X(P, \psi(a)) + (a - C_2)\psi'(a)x(P, \psi(a)) \; .$$

Let $\bar{P}(a)$ denote the dominant firm's profit-maximizing retail price for a given access charge a. In most reasonable cases, the dominant firm will set a higher retail price when the access

[223] This is just an instance of the general result that productive efficiency is desirable when there are enough tax instruments—see Diamond and Mirrlees (1971).

[224] This is adapted from section 7 of Laffont and Tirole (1994) 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), and Lapuerta and Tye (1999).

charge is higher (so $\bar{P}'(a) > 0$). This is the case because the more profit the dominant firm anticipates from selling access to its rivals, the less aggressively the dominant firm will compete with rivals at the retail level. The optimal access charge in this setting satisfies

$$a = \underbrace{C_2 + \sigma(\bar{P}(a) - C_1)}_{\text{ECPR charge}} - \frac{X\bar{P}'}{-z_a} \tag{103}$$

where σ is given in expression (100). Equation (103) reveals that the optimal access charge in this setting is *below* the level in the ECPR expression (99), which gives the optimal access charge in the setting where the dominant firm's retail price was fixed at $\bar{P}(a)$. The reason for the reduction in a when the dominant firm's retail price is unregulated is clear. A reduction in a here causes the retail price P to fall towards cost, which increases welfare.[225] (By contrast, in the Ramsey problem it was optimal to raise the access charge above (99)—see expression (102) 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 fixed costs.)

Another natural comparison is between a and the cost of access C_2. However, it is difficult to obtain clear-cut results about whether a is optimally set above or below cost. Either can be optimal. The special cases where the access charge should precisely equal cost include:

- Where the fringe has no ability to substitute away from the input and the demand functions X and x are linear.[226]
- When the dominant firm and fringe operate in separate retail markets, with no cross-price effects. (The profit-maximizing retail price $\bar{P}$ does not depend on a in this case. Also, since $\sigma = 0$, expression (103) implies that marginal cost pricing of access is optimal.)

A more interesting setting in which a cost-based access policy is optimal is where the fringe and the monopolist offer the same homogeneous product, i.e., where the retail market is potentially perfectly competitive. To see this, suppose that all consumers purchase from the supplier that offers the lowest retail price. If the access charge is a, the fringe will supply consumers whenever the incumbent offers a retail price greater than the fringe's cost, $\psi(a)$. Therefore, given a, the dominant firm has two options. First, it can preclude entry by the fringe by setting a retail price just below $\psi(a)$. Doing so ensures a profit of $\psi(a) - C_1$ per unit of retail output for the firm. Second, the dominant firm can choose not to operate in the retail market. If it does so (by, for example, choosing a retail price above $\psi(a)$), the dominant firm makes a profit of $(a - C_2)\psi'(a)$ per unit of retail output by selling access to the fringe. The dominant firm will choose to accommodate entry if and only if the latter profit margin exceeds the former, i.e., if

$$C_1 \geq \psi(a) - (a - C_2)\psi'(a) .$$

[225] A similar point is made in section III of Economides and White (1995). They show that when the downstream market is unregulated, it can be desirable to allow entry by an inefficient firm—something that is achieved by choosing an access charge below the ECPR level—if this causes retail prices to fall. In other words, it can be optimal to sacrifice some productive efficiency to reduce allocative inefficiency.

[226] See section 7 of Laffont and Tirole (1994) and Armstrong and Vickers (1998).

Since the right-hand side of this inequality is the total cost of a unit of fringe supply given the access charge a, the dominant firm will allow entry by the fringe if and only if supply by the fringe is less costly than supply by the dominant firm. Consequently, when the fringe firms are the least-cost suppliers, 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.[227]

More generally, when the monopolist has some market power in the retail market, the optimal access charge will equal cost only in knife-edge cases. Clear-cut results are difficult to obtain in this framework because the access charge 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 achieve allocative efficiency *given* P as discussed in section 5.1.1; and (iii) to pursue productive efficiency (which requires $a = C_2$) whenever there is a possibility for substituting away from the input. In general, task (i) and (iii) argue for an access charge no higher than cost. (When $a = C_2$ the dominant firm 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 (99). Because of these diverse, countervailing forces, it is not possible to give unequivocal guidance about the relationship between the access charge and the cost of providing access in unregulated retail markets.

5.1.5 Discussion

The primary benefits of setting access charges equal to the monopolist's costs are twofold. First, this policy is relatively simple to implement (provided the regulated firm's costs are readily estimated). In particular, no information about consumer demand or the characteristics of rivals is needed to calculate these charges (at least in the simple models presented above). Second, this is the only access pricing policy that ensures rivals offer their services at least cost. Pricing access above cost, as might be suggested by the ECPR policy for example, could induce an entrant to construct its own network, rather than purchase network services from the regulated firm, even though the latter entails lower social cost.

In simple terms, cost-based access charges are appropriate when access charges do not need to perform the role of correcting for distortions in the dominant firm's retail tariff. There are three main settings in which such a task may not be necessary:

1. First, if the regulated firm's retail tariff reflects its underlying costs, then no second-best corrective measures are needed. In such a setting, access charges should also reflect the relevant costs. In sum, a full and effective rebalancing of the regulated firm's tariff greatly simplifies the regulatory task, and allows access charges to focus on the task of ensuring productive efficiency.

2. Second, if there are distortions present in the regulated tariff, but the second-best corrections are made via another regulatory instrument (such as an output tax levied

[227] If industry costs are lower when the monopolist serves the market even when the fringe can purchase access at cost, 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 service at its cost.

on rivals), then access charges should reflect costs.

3. Third, when the input 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 competition with other suppliers.[228] 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 monopoly input supplier, 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 industry. 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.[229] The welfare increase can be particularly pronounced if the upstream monopolist can supply the retail service more efficiently than the other retailers.[230] Furthermore, downstream production by the upstream monopolist can deter some potential suppliers from entering the industry and thereby avoid duplicative fixed costs of production.[231]

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 input 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 retailing rivals. It can do this in at least two ways. First, the upstream producer may seek to raise the costs of downstream rivals by exaggerating its cost of supplying the essential input. If the upstream monopolist can convince the regulator that upstream production costs are high, the regulator may raise the price of the input, thereby increasing the operating costs of downstream competitors. By increasing the incentives of the upstream producer to exaggerate its operating costs in this manner, vertical integration can complicate the regulator's critical control problem.[232]

Second, the integrated firm may be able to raise its rivals' costs is by degrading the quality of the input it supplies or by imposing burdensome purchasing requirements on downstream

[228] See section 3.5.2 above for a discussion of the merits of allowing a regulated supplier to diversify into horizontally related markets.

[229] See Hinton, Zona, Schmalensee, and Taylor (1998) and Weisman and Williams (2001) for assessments of this effect in the U.S. telecommunications industry.

[230] See Lee and Hamilton (1999).

[231] See Vickers (1995a).

[232] 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.

producers, for example.[233] The regulator can affect the incentive an integrated supplier may have to raise the costs of its downstream rivals through the access charge. When the integrated producer enjoys a substantial profit margin on each unit of the input it sells to downstream producers, the integrated producer will sacrifice considerable upstream profit if it raises the costs of downstream rivals and thereby reduces their demand for the essential input. Therefore, the regulator may reduce any prevailing incentive to degrade quality raising the price of the essential input.[234] A detailed assessment of optimal regulatory policy in this regard remains to be conducted.

It should also be noted that a firm's participation in both upstream (input) and downstream (retail) markets 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 firm's prices. An aggregate restriction on overall price levels can admit a substantial increase in the price of one service (e.g., the essential upstream input that is sold to downstream competitors), as long as this increase is accompanied by a substantial decrease in the price of another service. Consequently, price cap regulation that applies to all of the prices set by a vertically integrated producer could allow the firm to exercise a price squeeze. A vertically integrated firm exercises a price squeeze when it charges its downstream competitors more for the essential input than it charges its downstream customers for a key retail service. As discussed in section 3.1.3 above, 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.[235]

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 being implemented.

6 Conclusions

This chapter has reviewed recent theoretical studies of the design of regulatory policy. We have focused on studies in which the regulated firm is assumed to have better information about its environment than does the regulator. The regulator's task in such settings often is to try to induce the regulated 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

[233] Economides (1998) examines a setting in which the incentives for raising rivals' costs in this manner are particularly pronounced. Also see Beard, Kaserman, and Mayo (2001), Reiffen, Schumann, and Ward (1998), section 4.5 of Laffont and Tirole (2000), Mandy (2000), and Mandy and Sappington (2003).

[234] Thus, one advantage of the ECPR policy discussed in the previous section, which might 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), and Sibley and Weisman (1998) for related analyses.

[235] See Laffont and Tirole (1996).

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 typically consider the implementation of a single regulatory policy that does not present the firm with an explicit choice among options. One interpretation of the non-Bayesian approach may be that regulatory plans that encompass options are "complicated", and therefore prohibitively costly to implement.[236] A second interpretation might be that the regulator has no information about the regulatory

[236] Ideally, the costs of complexity should be modeled explicitly, and the costs of more complicated regulatory plans should be weighed against their potential benefits.

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 negligible. 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?

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.[237] 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.

[237] 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).

References

Acton, J., and I. Vogelsang (1989): "Introduction to the Symposium on Price Cap Regulation," *Rand Journal of Economics*, 20(3), 369–372.

Anton, J., and P. Gertler (1988): "External Markets and Regulation," *Journal of Public Economics*, 37(2), 243–260.

Anton, J., J. Vander Weide, and N. Vettas (2002): "Entry Auctions and Strategic Behavior Under Cross-Market Price Constraints," *International Journal of Industrial Organization*, 20(5), 611–629.

Anton, J., and D. Yao (1987): "Second Sourcing and the Experience Curve: Price Competition in Defense Procurement," *Rand Journal of Economics*, 18(1), 57–76.

——— (1989): "Split Awards, Procurement and Innovation," *Rand Journal of Economics*, 20(4), 538–552.

——— (1992): "Coordination in Split Award Auctions," *Quarterly Journal of Economics*, 97(2), 681–707.

Armstrong, M. (1999): "Optimal Regulation with Unknown Demand and Cost Functions," *Journal of Economic Theory*, 84(2), 196–215.

——— (2000): "Optimal Multi-Object Auctions," *Review of Economic Studies*, 67(3), 455–481.

——— (2002): "The Theory of Access Pricing and Interconnection," in *Handbook of Telecommunications Economics: Volume I*, ed. by M. Cave, S. Majumdar, and I. Vogelsang. North-Holland, Amsterdam.

Armstrong, M., S. Cowan, and J. Vickers (1994): *Regulatory Reform: Economic Analysis and British Experience*. MIT Press, Cambridge, MA.

——— (1995): "Nonlinear Pricing and Price Cap Regulation," *Journal of Public Economics*, 58(1), 33–55.

Armstrong, M., R. Rees, and J. Vickers (1995): "Optimal Regulatory Lag Under Price Cap Regulation," *Revista Espanola De Economia*, 12, 93–116.

Armstrong, M., and J.-C. Rochet (1999): "Multi-Dimensional Screening: A User's Guide," *European Economic Review*, 43(4), 959–979.

Armstrong, M., and D. Sappington (2003): "Toward a Unified Treatment of Adverse Selection Problems," mimeo.

Armstrong, M., and J. Vickers (1991): "Welfare Effects of Price Discrimination by a Regulated Monopolist," *Rand Journal of Economics*, 22(4), 571–580.

——— (1993): "Price Discrimination, Competition and Regulation," *Journal of Industrial Economics*, 41(4), 335–360.

——— (1998): "The Access Pricing Problem With Deregulation: A Note," *Journal of Industrial Economics*, 46(1), 115–121.

——— (2000): "Multiproduct Price Regulation Under Asymmetric Information," *Journal of Industrial Economics*, 48(2), 137–160.

Auriol, E., and J.-J. Laffont (1992): "Regulation by Duopoly," *Journal of Economics and Management Strategy*, 1(3), 507–533.

Averch, H., and L. Johnson (1962): "Behavior of the Firm Under Regulatory Constraint," *American Economic Review*, 52(5), 1053–1069.

Avery, C., and T. Hendershott (2000): "Bundling and Optimal Auctions of Multiple Products," *Review of Economic Studies*, 67(3), 483–497.

Bailey, E., and R. Coleman (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.

——— (1988): "Regulation and Legislative Choice," *Rand Journal of Economics*, 29(3), 457–477.

——— (1989): "Design of Regulatory Mechanisms and Institutions," in *Handbook of Industrial Organization: Volume II*, ed. by R. Schmalensee, and R. Willig. North Holland, Amsterdam.

Baron, D., and D. Besanko (1984a): "Regulation and Information in a Continuing Relationship," *Information Economics and Policy*, 1(4), 267–302.

——— (1984b): "Regulation, Asymmetric Information, and Auditing," *Rand Journal of Economics*, 15(4), 447–470.

——— (1987a): "Commitment and Fairness in a Dynamic Regulatory Relationship," *Review of Economic Studies*, 54(3), 413–436.

——— (1987b): "Monitoring, Moral Hazard, Asymmetric Information, and Risk Sharing in Procurement Contracting," *Rand Journal of Economics*, 18(4), 509–532.

——— (1992): "Information, Control, and Organizational Structure," *Journal of Economics and Management Strategy*, 1(2), 237–275.

Baron, D., and R. Myerson (1982): "Regulating a Monopolist with Unknown Costs," *Econometrica*, 50(4), 911–930.

Baseman, K. (1981): "Open Entry and Cross-Subsidization in Regulated Markets," in *Studies in Public Regulation*, ed. by G. Fromm. The 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., E. BAILEY, AND R. WILLIG (1977): "Weak Invisible Hand Theorems on the Sustainability of Prices in a Multiproduct Monopoly," *American Economic Review*, 67(3), 350–365.

BAUMOL, W., AND A. KLEVORICK (1970): "Input Choices and Rate-of-Return Regulation: An Overview of the Discussion," *Bell Journal of Economics*, 1(2), 162–190.

BAUMOL, W., J. ORDOVER, AND R. WILLIG (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., J. PANZAR, AND R. WILLIG (1982): *Contestable Markets and the Theory of Industry Structure*. Harcourt Brace Jovanovich, Inc., New York.

BAUMOL, W., AND G. SIDAK (1994a): "The Pricing of Inputs Sold to Competitors," *Yale Journal on Regulation*, 11(1), 171–202.

——— (1994b): *Toward Competition in Local Telephony*. MIT Press, Cambridge, MA.

BEARD, R., D. KASERMAN, AND J. MAYO (2001): "Regulation, Vertical Integration, and Sabotage," *Journal of Industrial Economics*, 46(3), 319–333.

BERG, S., AND J. LYNCH (1992): "The Measurement and Encouragement of Telephone Service Quality," *Telecommunications Policy*, 16(3), 210–224.

BERNSTEIN, J., AND D. SAPPINGTON (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 *Analyzing the Impact of Regulatory Change in Public Utilities*, ed. by M. Crew. Lexington Books, Lexington, MA.

BESANKO, D., S. DONNENFELD, AND L. WHITE (1987): "Monopoly and Quality Distortion: Effects and Remedies," *Quarterly Journal of Economics*, 102(4), 743–767.

——— (1988): "The Multiproduct Firm, Quality Choice, and Regulation," *The Journal of Industrial Economics*, 36(4), 411–429.

BESANKO, D., AND D. SPULBER (1992): "Sequential-Equilibrium Investment by Regulated Firms," *Rand Journal of Economics*, 23(2), 153–170.

BESTER, H., AND R. STRAUSZ (2001): "Contracting with Imperfect Commitment and the Revelation Principle: The Single Agent Case," *Econometrica*, 69(4), 1077–1098.

BIAIS, B., AND E. PEROTTI (2002): "Machiavellian Privatization," *American Economic Review*, 92(1), 240–258.

BIGLAISER, G., AND A. MA (1995): "Regulating a Dominant Firm: Unknown Demand and Industry Structure," *Rand Journal of Economics*, 26(1), 1–19.

Biglaiser, G., and C.-T. A. Ma (1999): "Investment Incentives of a Regulated Dominant Firm," *Journal of Regulatory Economics*, 16(3).

Biglaiser, G., and M. Riordan (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.

Bolton, P., and M. Dewatripont (2002): *Introduction to Contract Theory*. MIT Press, Cambridge, MA, forthcoming.

Bolton, P., and J. Farrell (1990): "Decentralization, Duplication, and Delay," *Journal of Political Economy*, 98(4), 803–826.

Bradley, I., and C. Price (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.

——— (1984): "Socially Optimal Pricing with Rivalry and Economies of Scale," *Rand Journal of Economics*, 15(1), 127–134.

——— (1989): "Optimal Policies for Natural Monopolies," in *Handbook of Industrial Organization: Volume II*, ed. by R. Schmalensee, and R. Willig. North Holland, Amsterdam.

——— (1993): "A Regulatory Bargain for Diversified Enterprises," *International Journal of Industrial Organization*, 11(1), 1–20.

Braeutigam, R., and J. Panzar (1989): "Diversification Incentives Under 'Price-Based' and 'Cost-Based' Regulation," *Rand Journal of Economics*, 20(3), 373–391.

——— (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.

——— (1990): "Cross-Subsidization and Cost Misallocation by Regulated Monopolists," *Journal of Regulatory Economics*, 2(1), 37–52.

Brennan, T., and K. Palmer (1994): "Comparing the Costs and Benefits of Diversification by Regulated Firms," *Journal of Regulatory Economics*, 6(1), 115–136.

Caillaud, B. (1990): "Regulation, Competition, and Asymmetric Information," *Journal of Economic Theory*, 52(1), 87–110.

Caillaud, B., R. Guesnerie, and P. Rey (1992): "Noisy Observation in Adverse Selection Models," *Review of Economic Studies*, 59(3), 595–615.

Caillaud, B., R. Guesnerie, P. Rey, and J. Tirole (1988): "Government Intervention in Production and Incentives Theory: A Review of Recent Contributions," *Rand Journal of Economics*, 19(1), 1–26.

Celentani, M., and J.-J. Ganuza (2002): "Corruption and Competition in Procurement," *European Economic Review*, 46(7), 1273–1303.

Chang, Y.-M., and J. Warren (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.

——— (1995): "Revolving Doors and the Optimal Tolerance for Agency Collusion," *Rand Journal of Economics*, 26(3), 378–397.

Che, Y.-K., and I. Gale (1998): "Standard Auctions with Financially Constrained Bidders," *Review of Economic Studies*, 65(1), 1–22.

——— (2000): "The Optimal Mechanism for Selling to Budget-Constrained Consumers," *Journal of Economic Theory*, 92(2), 198–233.

Chiappori, P., I. Macho, P. Rey, and B. Salanié (1994): "Repeated Moral Hazard: The Role of Memory, Commitment and the Access to Credit Markets," *European Economic Review*, 38(8), 1527–1553.

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.

——— (1997b): "Tight Average Revenue Regulation Can be Worse than No Regulation," *Journal of Industrial Economics*, 45(1), 75–88.

——— (2002): "Optimal Risk Allocation for Regulated Monopolies and Consumers," *Journal of Public Economics*, forthcoming.

Crampes, C., and A. Hollander (1995): "Duopoly and Quality Standards," *European Economic Review*, 39, 71–82.

Cramton, P. (1997): "The FCC Spectrum Auctions: An Early Assessment," *Journal of Economics and Management Strategy*, 6, 431–495.

Cremér, J., and R. McLean (1985): "Optimal Selling Strategies Under Uncertainty for a Discriminating Monopolist When Demands are Interdependent," *Econometrica*, 53(2), 345–361.

Crew, M., and K. Crocker (1991): "Diversification and Regulated Monopoly," in *Competition and the Regulation of Utilities*, ed. by M. Crew. Kluwer Academic Publishers, Boston, MA.

Crew, M., and P. Kleindorfer (2002): "Regulatory Economics: Twenty Years of Progress?," *Journal of Regulatory Economics*, 21(1), 5–22.

Cripps, M., and N. Ireland (1994): "The Design of Auctions and Tenders with Quality Thresholds: The Symmetric Case," *Economic Journal*, 104(423), 316–326.

Crémer, J., F. Khalil, and J.-C. Rochet (1998a): "Contracts and Productive Information Gathering," *Games and Economic Behavior*, 23(2), 174–193.

——— (1998b): "Strategic Information Gathering Before a Contract is Offered," *Journal of Economic Theory*, 81(1), 163–200.

Curien, N., B. Jullien, and P. Rey (1998): "Pricing Regulation Under Bypass Competition," *Rand Journal of Economics*, 29(2), 259–279.

Da Rocha, J., and M. A. de Frutos (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.

——— (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., and K. Spier (1994): "Designing a Private Industry: Government Auctions with Endogenous Market Structure," *Journal of Public Economics*, 53(1), 127–147.

Demsetz, H. (1968): "Why Regulate Utilities?," *Journal of Law and Economics*, 11(1), 55–65.

Demski, J., and D. Sappington (1984): "Optimal Incentive Contracts With Multiple Agents," *Journal of Economic Theory*, 33(1), 152–171.

——— (1987): "Hierarchical Regulatory Control," *Rand Journal of Economics*, 18(3), 369–383.

Demski, J., D. Sappington, and P. Spiller (1987): "Managing Supplier Switching," *Rand Journal of Economics*, 18(1), 77–97.

——— (1988): "Incentive Schemes With Multiple Agents and Bankruptcy Constraints," *Journal of Economic Theory*, 44(1), 156–167.

DIAMOND, P., AND J. MIRRLEES (1971): "Optimal Taxation and Public Production: I - Production Efficiency," *American Economic Review*, 61(1), 8–27.

ECKEL, C., AND N. LUTZ (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., AND L. WHITE (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., AND D. SAPPINGTON (1995): "Toward a Benchmark for Optimal Prudency Policy," *Journal of Regulatory Economics*, 7(2), 111–131.

FINSINGER, J., AND I. VOGELSANG (1981): "Alternative Institutional Frameworks for Price Incentive Mechanisms," *Kyklos*, 34(3), 338–404.

——— (1982): "Performance Indices for Public Enterprises," in *Public Enterprise in Less Developed Countries*, ed. by L. Jones. Cambridge University Press, Cambridge.

——— (1985): "Strategic Management Behavior Under Reward Structures in a Planned Economy," *Quarterly Journal of Economics*, 100(1), 263–270.

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., R. GUESNERIE, AND J. TIROLE (1985): "Planning Under Incomplete Information and the Ratchet Effect," *Review of Economic Studies*, 52(2), 173–191.

FUDENBERG, D., AND J. TIROLE (1990): "Moral Hazard and Renegotiation in Agency Contracts," *Econometrica*, 58(6), 1279–1320.

——— (1991): *Game Theory*. MIT Press, Cambridge, MA.

FULLERTON, R., AND P. MCAFEE (1999): "Auctioning Entry Into Tournaments," *Journal of Political Economy*, 107(3), 573–605.

GASMI, F., M. IVALDI, AND J.-J. LAFFONT (1994): "Rent Extraction and Incentives for Efficiency in Recent Regulatory Proposals," *Journal of Regulatory Economics*, 6(2), 151–176.

GILBERT, R., AND D. NEWBERY (1994): “The Dynamic Efficiency of Regulatory Constitutions,” *Rand Journal of Economics*, 25(4), 538–554.

GILBERT, R., AND M. RIORDAN (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.

GRAVELLE, H. (1985): “Reward Structures in a Planned Economy: Some Difficulties,” *Quarterly Journal of Economics*, 100(1), 271–278.

GREENWALD, B. (1984): “Rate Base Selection and the Structure of Regulation,” *Rand Journal of Economics*, 15(1), 85–95.

GUESNERIE, R., AND J.-J. LAFFONT (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.

HAGERMAN, J. (1990): “Regulation by Price Adjustment,” *Rand Journal of Economics*, 21(1), 72–82.

HARRIS, M., AND R. TOWNSEND (1981): “Resource Allocation Under Asymmetric Information,” *Econometrica*, 49(1), 33–64.

HILLMAN, J., AND R. BRAEUTIGAM (1989): *Price Level Regulation for Diversified Public Utilities*. Boston, MA.

HINTON, P., J. D. ZONA, R. SCHMALENSEE, AND W. TAYLOR (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., AND P. MILGROM (1987): “Aggregation and Linearity in the Provision of Intertemporal Incentives,” *Econometrica*, 55(2), 303–328.

IOSSA, E. (1999): “Informative Externalities and Pricing in Regulated Multiproduct Industries,” *Journal of Industrial Economics*, 47(2), 195–220.

IOSSA, E., AND F. STROFFOLINI (2002): “Price Cap Regulation and Information Acquisition,” *International Journal of Industrial Organization*, 20(7), 1013–1036.

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.

JULLIEN, B. (2000): “Participation Constraints in Adverse Selection Models,” *Journal of Economic Theory*, 93(1), 1–47.

Kang, J., D. Weisman, and M. Zhang (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., and C.-Y. Jung (1995): "Regulating a Multi-Product Monopolist," *Journal of Regulatory Economics*, 8(3), 299–307.

Kjerstad, E., and S. Vagstad (2000): "Procurement Auctions with Entry of Bidders," *International Journal of Industrial Organization*, 18(8), 1243–1257.

Kofman, F., and J. Lawarrée (1993): "Collusion in Hierarchical Agency," *Econometrica*, 61(3), 629–656.

Kolbe, L., and W. Tye (1991): "The Duquesne Opinion: How Much 'Hope' is There for Investors in Regulated Firms," *Yale Journal of Regulation*, 8(1), 113–157.

Kridel, D., D. Sappington, and D. Weisman (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 *Price Caps and Incentive Regulation in Telecommunications*, ed. by M. Einhorn. Kluwer Academic Publishers, Boston.

——— (1993): "Implementing Price Caps in Telecommunications," *Journal of Policy Analysis and Management*, 12(4), 726–752.

Laffont, J.-J., and D. Martimort (1997): "Collusion Under Asymmetric Information," *Econometrica*, 65(4), 875–912.

——— (1999): "Separation of Regulators Against Collusive Behavior," *Rand Journal of Economics*, 30(2), 232–262.

——— (2002): *The Theory of Incentives: The Principal-Agent Model*. Princeton University Press, Princeton, NJ.

Laffont, J.-J., and J. Tirole (1986): "Using Cost Observation to Regulate Firms," *Journal of Political Economy*, 94(3), 614–641.

——— (1987): "Auctioning Incentive Contracts," *Journal of Political Economy*, 95(5), 921–937.

——— (1988a): "The Dynamics of Incentive Contracts," *Econometrica*, 56(5), 1153–1176.

——— (1988b): "Repeated Auctions of Incentive Contracts, Investment, and Bidding Parity with an Application to Takeovers," *Rand Journal of Economics*, 19(4), 516–537.

——— (1990a): "Adverse Selection and Renegotiation in Procurement," *Review of Economic Studies*, 57(4), 597–626.

——— (1990b): "Optimal Bypass and Cream Skimming," *American Economic Review*, 80(4), 1041–1051.

——— (1990c): "The Politics of Government Decision-Making: Regulatory Institutions," *Journal of Law, Economics, and Organization*, 6(1), 1–32.

——— (1991a): "Auction Design and Favoritism," *International Journal of Industrial Organization*, 9(1), 9–42.

——— (1991b): "The Politics of Government Decision-Making: A Theory of Regulatory Capture," *Quarterly Journal of Economics*, 106(4), 1089–1127.

——— (1991c): "Privatization and Incentives," *Journal of Law, Economics, and Organization*, 7(3), 84–105.

——— (1993a): "Cartelization by Regulation," *Journal of Regulatory Economics*, 5(2), 111–130.

——— (1993b): *A Theory of Incentives in Procurement and Regulation*. MIT Press, Cambridge, MA.

——— (1994): "Access Pricing and Competition," *European Economic Review*, 38(9), 1673–1710.

——— (1996): "Creating Competition Through Interconnection: Theory and Practice," *Journal of Regulatory Economics*, 10(3), 227–256.

——— (2000): *Competition in Telecommunications*. MIT Press, Cambridge, MA.

Lapuerta, C., and W. Tye (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.

——— (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.

——— (1997b): "A Note on Regulating Oligopolistic Industries," *Journal of Regulatory Economics*, 12(1), 91–97.

LEE, S.-H., AND J. HAMILTON (1999): "Using Market Structure to Regulate a Vertically Integrated Monopolist," *Journal of Regulatory Economics*, 15(3), 223–248.

LEHMAN, D., AND D. WEISMAN (2000): "The Political Economy of Price Cap Regulation," *Review of Industrial Organization*, 16(4), 343–356.

LEVY, B., AND P. SPILLER (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., AND D. SAPPINGTON (1988a): "Regulating a Monopolist with Unknown Demand," *American Economic Review*, 78(5), 986–998.

——— (1988b): "Regulating a Monopolist with Unknown Demand and Cost Functions," *Rand Journal of Economics*, 18(3), 438–457.

——— (1989a): "Countervailing Incentives in Agency Problems," *Journal of Economic Theory*, 49(2), 294–313.

——— (1989b): "Inflexible Rules in Incentive Problems," *American Economic Review*, 79(1), 69–84.

——— (1989c): "An Informational Effect When Regulated Firms Enter Unregulated Markets," *Journal of Regulatory Economics*, 1(1), 35–46.

——— (1989d): "Regulatory Options and Price Cap Regulation," *Rand Journal of Economics*, 20(3), 405–416.

——— (1990): "Sequential Regulatory Oversight," *Journal of Regulatory Economics*, 2(4), 327–348.

——— (1991a): "Incentives for Monitoring Quality," *Rand Journal of Economics*, 22(3), 370–384.

——— (1991b): "Oversight of Long-Term Investment by Short-Lived Regulators," *International Economic Review*, 32(3), 579–600.

——— (1992): "Incentives for Conservation and Quality Improvement by Public Utilities," *American Economic Review*, 82(5), 1321–1340.

——— (1997): "Information Management in Incentive Problems," *Journal of Political Economy*, 105(4), 796–821.

——— (1999): "Access Pricing With Unregulated Downstream Competition," *Information Economics and Policy*, 11(1), 73–100.

——— (2000): "Motivating Wealth-Constrained Actors," *American Economic Review*, 90(4), 944–960.

LEWIS, T., AND H. YILDIRIM (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., AND W. MAGAT (1979): "A Decentralized Method for Utility Regulation," *Journal of Law and Economics*, 22(2), 399–404.

LUTON, R., AND P. MCAFEE (1986): "Sequential Procurement Auctions," *Journal of Public Economics*, 31(2), 181–195.

LYNCH, J., T. BUZAS, AND S. BERG (1994): "Regulatory Measurement and Evaluation of Telephone Service Quality," *Management Sceience*, 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.

——— (1992): "Regulation With 20-20 Hindsight: Least-Cost Rules and Variable Costs," *Journal of Industrial Economics*, 40(3), 277–289.

——— (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., J. MOORE, AND S. TURNBULL (1988): "Stopping Agents From 'Cheating'," *Journal of Economic Theory*, 46(2), 335–372.

MAGGI, G., AND A. RODRIGUEZ-CLARE (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., AND D. SAPPINGTON (2003): "Incentives for Sabotage in Vertically-Related Industries," mimeo, University of Missouri.

MANELLI, A., AND D. VINCENT (1995): "Optimal Procurement Mechanisms," *Econometrica*, 63(3), 591–620.

MANKIW, G., AND M. WHINSTON (1986): "Free Entry and Social Inefficiency," *Rand Journal of Economics*, 17(1), 48–58.

MANSELL, R., AND J. CHURCH (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, P., AND J. MCMILLAN (1996): "Analyzing the Airwaves Auction," *Journal of Economic Perspectives*, 10(1), 159–175.

MCAFEE, R. P., AND J. MCMILLAN (1987a): "Auctions and Bidding," *Journal of Economic Literature*, 25(2), 699–738.

——— (1987b): "Competition for Agency Contracts," *Rand Journal of Economics*, 18(2), 296–307.

MCGUIRE, T., AND M. RIORDAN (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., AND J. VICKERS (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.

MOOKHERJEE, D. (1984): "Optimal Incentive Schemes with Many Agents," *Review of Economic Studies*, 51(3), 433–446.

MUSSA, M., AND S. ROSEN (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., AND J. STIGLITZ (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. (2000): *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.

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.

——— (1985): "Repeated Principal-Agent Relationships With Discounting," *Econometrica*, 53(5), 1173–1198.

Ramakrishnan, R. T. S., and A. V. Thakor (1991): "Cooperation versus Competition in Agency," *Journal of Law, Economics and Organization*, 7(2), 248–283.

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., L. Schumann, and M. Ward (1998): "Discriminatory Dealing with Downstream Competitors: Evidence From the Cellular Industry," *Federal Trade Commission Working Paper*.

Rey, P., and B. Salanié (1996): "On the Value of Commitment with Asymmetric Information," *Econometrica*, 64(6), 1395–1414.

Rey, P., and J. Tirole (2003): "A Primer on Foreclosure," in *Handbook of Industrial Organization: Volume III*, ed. by M. Armstrong, and R. Porter. North-Holland, Amsterdam, forthcoming.

Riordan, M. (1984): "On Delegating Price Authority to a Regulated Firm," *Rand Journal of Economics*, 15(1), 108–115.

——— (1996): "Contracting With Qualified Suppliers," *International Economic Review*, 37(1), 115–128.

——— (2002): "Universal Residential Telephone Service," in *Handbook of Telecommunications Economics: Volume I*, ed. by M. Cave, S. Majumdar, and I. Vogelsang. North-Holland, Amsterdam.

Riordan, M., and D. Sappington (1987a): "Awarding Monopoly Franchises," *American Economic Review*, 77(3), 375–387.

——— (1987b): "Information, Incentives, and Organizational Mode," *Quarterly Journal of Economics*, 102(2), 243–263.

——— (1988): "Optimal Contracts with Public Ex Post Information," *Journal of Economic Theory*, 45(1), 189–199.

——— (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., AND L. STOLE (2003): "The Economics of Multidimensional Screening," in *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, ed. by M. Dewatripont, L. P. Hansen, and S. Turnovsky. Cambridge University Press, Cambridge.

ROGERSON, W. (1985): "Repeated Moral Hazard," *Econometrica*, 53(1), 69–76.

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.

——— (2000): "Auctions and Regulation: Reengineering of Regulatory Mechanisms," *Journal of Regulatory Economics*, 17(3), 195–204.

SALANT, D., AND G. WOROCH (1992): "Trigger Price Regulation," *Rand Journal of Economics*, 23(1), 29–51.

SAPPINGTON, D. (1980): "Strategic Firm Behavior Under a Dynamic Regulatory Adjustment Process," *Bell Journal of Economics*, 11(1), 360–372.

——— (1982): "Optimal Regulation of Research and Development Under Imperfect Information," *Bell Journal of Economics*, 13(2), 354–368.

——— (1983): "Limited Liability Contracts Between Principal and Agent," *Journal of Economic Theory*, 29(1), 1–21.

——— (1986): "Commitment to Regulatory Bureaucracy," *Information Economics and Policy*, 2(4), 243–258.

——— (1994): "Designing Incentive Regulation," *Review of Industrial Organization*, 9(3), 245–272.

——— (2002): "Price Regulation and Incentives," in *Handbook of Telecommunications Economics: Volume I*, ed. by M. Cave, S. Majumdar, and I. Vogelsang. North Holland, Amsterdam.

——— (2003): "Regulating Horizontal Diversification," *International Journal of Industrial Organization*, 21(3), 291–315.

SAPPINGTON, D., AND D. SIBLEY (1988): "Regulating Without Cost Information: The Incremental Surplus Subsidy Scheme," *International Economic Review*, 29(2), 297–306.

——— (1990): "Regulating Without Cost Information: Further Observations," *International Economic Review*, 31(4), 1027–1029.

——— (1992): "Strategic Nonlinear Pricing Under Price Cap Regulation," *Rand Journal of Economics*, 23(1), 1–19.

——— (1993): "Regulatory Incentive Policies and Abuse," *Journal of Regulatory Economics*, 5(2), 131–141.

Sappington, D., and D. Weisman (1996a): *Designing Incentive Regulation for the Telecommunications Industry.* MIT Press, Cambridge, MA.

——— (1996b): "Revenue Sharing in Incentive Regulation Plans," *Information Economics and Policy*, 8(3), 229–248.

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., and D. Weisman (1998): "Raising Rivals' Costs: The Entry of an Upstream Monopolist Into Downstream Markets," *Information Economics and Policy*, 10(4), 451–470.

Sidak, G., and D. Spulber (1997): *Deregulatory Takings and the Regulatory Contract.* Cambridge University Press, Cambridge.

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., AND D. SPULBER (1994): "The Capital Structure of Regulated Firms," *Rand Journal of Economics*, 25(3), 424–440.

——— (1997): "Capital Structure With Countervailing Incentives," *Rand Journal of Economics*, 28(1), 1–24.

SPILLER, P. (1990): "Polititians, 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.

TAM, M.-Y. (1981): "Reward Structures in a Planned Economy: Some Difficulties," *Quarterly Journal of Economics*, 96(1), 111–128.

TANGERÅS, T. (2002): "Collusion-Proof Yardstick Competition," *Journal of Public Economics*, 83(2), 231–254.

TIROLE, J. (1986a): "Hierarchies and Bureaucracies: On the Role of Collusion in Organizations," *Journal of Law, Economics and Organization*, 2(2), 181–214.

——— (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.

——— (1995a): "Competition and Regulation in Vertically Related Markets," *Review of Economic Studies*, 62(1), 1–17.

——— (1995b): "Concepts of Competition," *Oxford Economic Papers*, 47(1), 1–23.

VOGELSANG, I. (1983): "Effort Rewarding Incentive Mechanisms for Public Enterprise Managers," *International Journal of Industrial Organization*, 1(2), 253–273.

——— (1988): "A Little Paradox in the Design of Regulatory Mechanisms," *International Economic Review*, 29(3), 467–476.

——— (1989): "Price Cap Regulation of Telecommunications Services: A Long-Run Approach," in *Deregulation and Diversification of Utilities*, ed. by M. Crew. Kluwer Academic Publishers, Boston.

——— (1990): "Optional Two-Part Tariffs Constrained by Price Caps," *Economics Letters*, 33(3), 287–292.

——— (2002): "Incentive Regulation and Competition in Public Utility Markets: A 20-Year Perspective," *Journal of Regulatory Economics*, 22(1), 5–28.

VOGELSANG, I., AND J. FINSINGER (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.

——— (1994): "Why More May be Less under Price-Cap Regulation," *Journal of Regulatory Economics*, 6(4), 339–362.

——— (1995): "Regulation and the Vertically Integrated Firm: The Case of RBOC Entry Into InterLATA Long Distance," *Journal of Regulatory Economics*, 8(3), 249–266.

——— (1998): "The Incentive to Discriminate by a Vertically-Integrated Firm: A Reply," *Journal of Regulatory Economics*, 14(1), 87–91.

WEISMAN, D., AND M. WILLIAMS (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.

——— (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 *Issues in Public Utility Regulation*, ed. by H. Trebing. Michigan State University Press, East Lansing, MI.

WILSON, R. (1979): "Auctions of Shares," *Quarterly Journal of Economics*, 93(4), 675–689.

ZUPAN, M. (1989a): "Cable Franchise Renewals: Do Incumbent Firms Behave Opportunistically?," *Rand Journal of Economics*, 20(4), 473–482.

——— (1989b): "The Efficacy of Franchise Bidding Schemes in the Case of CATV: Some Systematic Evidence," *Journal of Law and Economics*, 32(2), 401–456.