Regulating a Monopolist with Unverifiable Quality: The Effect of Firm Objectives and Consumer Responsiveness on Output*

Kevin E. Pflum†

I revisit a regulator's problem when a monopolist has superior knowledge of either cost or demand which is responsive to unverifiable quality. Many markets possessing unverifiable quality (e.g., health care services and education) often contain a mix of for-profit and not-for profit firms, therefore the firm's objective is modeled as a combination of profit- and output-maximization. In contrast to the earlier literature, the firm's ability to extract information rents in combination with consumers' responsiveness to quality may result in either an under- or over-supply of the good relative to first-best in every state. At the extreme, however, the firm's informational advantage may be completely attenuated when consumer's price elasticity is eliminated and the firm is an output-maximizer. The findings provide new insights into how the strategic response of the firm and consumers to the regulator's payment policy may lead to market distortions and what form these distortions take.
JEL Codes: H42, L20, L31, L50
Key Words: regulating quality, firm objectives, adverse selection, moral hazard

In early 2010 the U.S. Congress passed the Patient Protection and Affordable Care Act (PPACA) in an effort to increase access, lower costs, and improve quality for health care. To this end, a provision of the bill sets aside funds for the Centers for Medicare and Medicaid Services (CMS) to research, develop, and test new payment and delivery arrangements for health care providers. There of course exists a large body of literature examining the regulation of public utilities to draw upon when designing new payment mechanisms; however, the regulatory approaches taken for public utilities may not be completely transferable to health markets. For example, in telecommunications quality can be partially identified by quantitative measures such as the time to connect or the drop call ratio, and with public water quality can be identified by the quantitative measure of contaminant parts per million. In contrast, in health markets quality may refer to treatment techniques, intensities, or technological sophistication that cannot be easily defined or measured, even if observable by a regulator. Consequently, minimum service quality regulation may be impractical or undesirable for health services. Moreover, rate-of-return regulation may also be impractical in such markets where the cost of providing the good or service to an individual is easily disguised due to consumer heterogeneity and economies of scope.

Similar limitations to regulating quality also arise in other markets such as public and higher education. For instance, the 2008 Charter School Renewal Quality Review Handbook for the Oakland Unified School District provides an itemized list of characteristics it uses to measure the

* I am grateful to Paul J. Healy, Dan Levin, Lixin Ye, Huanxing Yang, Michael Sinkey, Paula Cordero Salas, and seminar participants at the Ohio State University, Kent State University, Bowling Green State University, Miami University, and the University of Alabama as well as conference participants at the 2012 ASSA winter meeting for helpful discussions and comments.
† Address: Department of Economics, Finance and Legal Studies, Box 870224, Tuscaloosa, AL, 35406, kpflum@cba.ua.edu.

quality of schools. The difficulty in quantifying quality levels, however, is reflected in the following statement from the handbook report (emphasis added):

> It is also imperative that everyone recognizes that there are many ways in which a school's program for improving student outcomes can merit a particular evaluation and that awarding levels is a matter of *informed professional judgment* and not simply a technical process.

Given the unique challenges quality introduces to a regulatory environment, the purpose of this study is to take a new look at the design of regulatory policy for a firm such as a hospital that can manipulate demand through its choice of unverifiable quality. Traditionally the mechanism design literature on regulating firms with unverifiable quality has assumed that demand is inelastic to quality, or has modeled quality in an environment of observable output and cost.[1] When demand can be manipulated by the firm's choice of quality the firm really has only one choice variable: the quantity of the good that it wants to sell. Reflecting this inseparability, we transform the firm and regulator's problems to make this relationship explicit. That is, despite the presence of quality, the firm has only one choice variable: the quantity of output. This relationship between quantity and quality along with the lack of information on which to contract produces substantially different outcomes than those found in the previous literature.

Markets in which consumer demand is affected by quality—such as health services, insurance, and education—generally contain a mix of for-profit and not-for-profit firms.[2] Therefore, the firm's objective function is represented as a weighted sum of profit and community-benefit. This allows us to analyze how the firm's objectives affect the regulator's problem. The objective function admits as special cases a pure profit-maximizer and a pure output-maximizer for easy comparison with the previous literature. Additionally, the consumers' access to the good is altered by considering a scenario where consumers are responsible for paying for their consumption directly and a scenario where the regulator pays on behalf of consumers using funds raised via taxation. Adopting the terminology of Caillaud, Guesnerie, and Tirole (1988), we refer to the good in the former scenario as a *marketed* good, and in the latter as *nonmarketed*. The marketed good represents the classical regulatory environment generally associated with public utilities.[3] The nonmarketed good represents the regulatory environment most often attributed to markets for health care,[4] but is relevant for any market in which the government is responsible for the provision of the good. Examples in the U.S. where the government is responsible for the payment, but not the production of the good, include public education provided by charter schools, voter registration services, military contracting, and of course, healthcare through Medicare, Medicaid, and the State Children's Health Insurance Program (SCHIP).

As has long been understood in Bayesian mechanism design problems, the firm is often able to extract an information rent in the presence of asymmetric information; however, the contribution to the distortion away from the social optimum caused by the interaction of the firm's and consumers'

[1] For example, in their treatment of quality Laffont and Tirole (1993) consider models where quality replaces effort and models having both quality and effort but where quality can be substituted with a linear transformation of verifiable quantity eliminating the problem of unverifiability.

[2] For example, of the 4,897 registered community hospitals in the United States, 873 are for-profit while 2,913 are non-profit (AHA fast facts on US hospitals: http://www.aha.org/aha/resource-center/Statistics-and-Studies/fast-facts.html), and twelve states permit for-profit corporations to operate charter schools (National Education Association: http://www.corpwatch.org/article.php?id=886).

[3] Seminal works include Myerson (1979), Baron and Besanko (1984), Laffont and Tirole (1986), Sappington (1982), Riordan (1984) and Lewis and Sappington (1988)

[4] See Ellis and McGuire (1986), Ma (1994), and Chalkley and Malcomson (1998b).

best-response to the regulator's payment rules has not been thoroughly explored.[5] For example, the key finding of Lewis and Sappington (1988) is that, when there is asymmetric information about the demand state, the firm is unable to benefit from its informational advantage and the regulator can induce the socially optimal output.[6] This result is in sharp contrast to when the asymmetric information is with respect to the firm's cost. In this case, Baron and Myerson (1982) show that the firm is able to extract an information rent resulting in an output that is always distorted down from the socially optimal level.

The contrasting findings of these two studies highlight the fact that asymmetric information is not a sufficient condition for a distortion away from the social optimum. Moreover, moral hazard does not necessarily lead to distortions either. Caillaud et al. (1988) first note this with an example of hidden effort and unobserved firm cost. There is no distortion away from the social optimum in this case because the agent internalizes all of the gains from exerting effort. The agent consequently exerts effort up to the point that the marginal benefit of additional effort is equal to the marginal cost, which is exactly the social optimum. When the agent is reimbursed based on an observable cost, then its cost savings from exerting effort are not internalized and the agent's incentives are altered in such a way that the agent's choice of effort is distorted away from the social optimum. Examples of such distortions abound in Laffont and Tirole (1993) who utilize a framework of compensation based on observable costs.[7]

Unlike firm effort, unverifiable quality represents a dimension of moral hazard that affects consumers. Consequently, quality can result in a distortion away from the social optimum without any adverse selection (Spence, 1975; Baron, 1981). This distortion occurs because of a conflict in the response by the firm and consumers to the regulator's policy instrument and the fact that the firm can manipulate consumer demand through its choice of quality. On the other hand, if consumer demand is price inelastic, then the regulator's policy instrument only affects the firm and there will be no distortion in output (Ma, 1994).[8] These differences highlight the importance of accounting for both the consumers' and firm's strategic response to the contract instruments.

This paper generates two principal findings. First, in a regulatory environment containing a marketed good, the results of the model show that one cannot predict *a priori* what form the output distortion will take even for a pure profit-maximizing firm. Depending on the relative price- and quality-elasticities of demand and cost, a firm's rents may either increase or decrease with a change in the unit payment resulting in a distortion of output that results in either an over- *or* under-supply relative to the first-best level. This finding is in marked contrast to the familiar under-supply relative to first-best outcome obtained in similar models that either do not include quality (implicitly setting the quality-elasticity of demand to zero), explicitly set quality elasticity to zero, or have observable firm costs. When the good is nonmarketed, however, the distortion away from the social optimum is similar to the earlier literature because the firm's information rents are strictly increasing with the unit payment forcing the regulator to set a lower price in order to limit the firm's rents.

[5] Armstrong and Sappington (2004) have taken a first step in providing a synthesis of the regulatory problem. Their focus, however, is on adverse selection and how it affects the regulator's payment rules and the market outcomes and do not consider the additional problems created by moral hazard.

[6] Aguirre and Beitia (2004) modify the model of Lewis and Sappington (1988) by making the consumer responsible for the unit payment, and the regulator responsible for the transfer payment. When the payment is split, the regulator has a strict preference for unit payments over the fixed transfer because of the deadweight loss attributed to raising public funds. This preference for one payment over the other prevents the regulator from achieving the socially optimal outcome and, as such, is a step backwards for the regulator, who can achieve the socially optimal outcome if either party is responsible for both payments.

[7] See also Laffont and Martimort (2002) and Armstrong and Sappington (2004) for references to related models.

[8] Ma (1994) finds that there will be distortions in quality when the firm can dump costly patients, or "cream-skim" to treat the least costly.

Our second finding is that the firm's informational advantage does not necessarily result in an output distortion. Not surprisingly, we find that the second-best payment policy takes on a very different form depending on the objective of the firm and the consumers' access to the good. Somewhat surprisingly, however, we find that the regulator can completely attenuate the firm's informational advantage when the good is nonmarketed and the firm's preference for community benefit over profit is sufficiently strong that it acts as a pure output-maximizer. Note that the firm is not a perfect agent of the regulator as in Ellis and McGuire (1986), but rather the nature of the firm's objective allows the regulator to more precisely control the firm's choice of output with the available policy instruments. This finding indicates that programs like Medicare, which use the same payment rules for firms exhibiting very different objectives will result in sub-optimal outcomes.[9] To eliminate the firm's informational advantage, the regulator faces a trade-off in reducing the firm's informational advantage and incurring the social deadweight loss from raising public funds. Thus, the analysis can also be thought of as providing insight into when it is beneficial to utilize public funds to provide access to a good or service and when it is not; e.g., a single-payer versus market oriented healthcare payment system.

We utilize the standard techniques of Bayesian mechanism design so the mechanics of the paper are the same or similar to those found in many papers studying the regulation of firms (e.g., Baron and Myerson (1982), Laffont and Tirole (1986), Caillaud et al. (1988), and Lewis and Sappington (1992)). The information framework of our model is closest to Lewis and Sappington (1992) who study the design of incentive programs to induce public utilities to provide a basic service with enhancements. Lewis and Sappington (1992) partially analyze the optimal contracts when quality enhancement is observable, the quantity consumed is observable, and when neither are observable. They do not consider other incentive regimes such as nonmarketed goods or mixed-objectives firms, nor do they fully characterize the distortions from the social optimum.

Finally, this paper also relates to the literature that looks at optimal cost sharing rules for a firm having a partially altruistic objective. Ellis and McGuire (1986) first consider a firm with similar preferences as the regulator. The authors show that when the preferences of the firm are the same as the regulator's, the appropriate payment rule will induce the firm to produce the socially preferred level of quality. If the firm has less of a preference for consumer benefits, however, then the regulator must subsidize the firm in order to induce the socially preferred level of quality. Chalkley and Malcomson (1998a) introduce the possibility of cost-reducing effort by the firm and find that the regulator can induce the socially optimal level of effort, similar to Caillaud et al. (1988), but at a sub-optimal level of quality. Finally, Jack (2005) introduces uncertainty in the degree of altruism for a firm and derives the optimal cost-sharing rules when quality is observable, unobservable, and when the firm has an unknown degree of altruism.

There are three ways in which these analyses differ from this paper. First, the firm's costs are not observable in the current model. Second, demand is inelastic to quality in these models, which by itself simplifies the regulator's problem, but also requires that any incentive to produce quality is completely motivated by altruism. Because higher quality can attract more consumers, the profit motive is sufficient to produce high quality and an added altruistic objective enhances the incentive to produce quality. Lastly, in this paper's model the firm is restricted to a zero profit constraint substantially altering the analysis when the firm has a sufficiently strong preference for community benefit that causes it to bump up against the constraint.

[9]In a recent study Landon et al. (2006) examined the quality of care for myocardial infarction, congestive heart failure, and pneumonia provided by for-profit and nonprofit hospitals and, consistent with the findings of this paper, find that patients were more likely to receive higher quality care in a nonprofit hospital reflecting the differences in outcome that result from the same payment scheme.

The derivation of the our findings progresses as follows. In Section I the primary model is developed, including the derivation of the *quality-adjusted* cost and demand functions. In section II we establish the base-line, socially-optimal contract. In Section III we analyze the regulator's problem when it chooses to have consumers pay directly for the good or service, and in Section IV we analyze the regulator's problem when it pays for the good on behalf of consumers.[10] Finally, in Section V we summarize the findings of the paper and provide some direction for future research. Unless otherwise stated, all proofs appear in the Appendix.

I. The General Model

Consumers

Consider a market environment where there is a single firm supplying a good or service at some level of quality $q \in \mathbb{R}_+$. Quality is observable by consumers but is not verifiable, so it cannot be directly contracted upon.[11] We proceed by assuming that the firm has superior knowledge of its cost and in an Appendix show that the results are robust to asymmetric information about consumer demand. Demand for the firm's good, $x(p, q)$, is a function of the price p and the quality, q, of the good.[12,13] Consumer demand can also represent a residual demand function for an imperfectly competitive market as with the market for hospital services, which includes substantial spacial competition. Under this interpretation, $x(p, q)$ represents the firm's residual demand given all other firms supply some equilibrium level of quality and the regulator is assumed to take the number of firms in the market as given.[14] Demand is $\mathcal{C}^2$, increasing and concave in q,[15] and decreasing in p. The regulator cannot observe the quantity of the good or services consumed, otherwise knowing the price, the regulator will know the quality level.

Gross consumer benefit is represented by the function $B(x, q)$.[16] B may reflect the consumers' direct value of consumption (i.e., $B(x, q) = \int_0^x P(\tilde{x}, q) d\tilde{x}$ where P is inverse demand) or B may reflect the regulator's valuation of consumption in the presence of social externalities as are common with services such as health care and education. Gross consumer benefit is increasing and concave in the quantity and quality; $B_x > 0$, $B_q > 0$, $B_{xx} \leq 0$, and $B_{qq} \leq 0$.[17] Moreover, marginal consumer benefit is weakly increasing in quality, $B_{xq} \geq 0$, reflecting the complementarity of quality and output.

The Firm

The cost to the firm of producing quantity x at quality q is given by the function $c(x, q; \theta)$, which is parameterized by the cost-state $\theta \in \Theta$, where Θ is some closed interval of the real number

[10] The robustness of the results to asymmetric demand information are established in an Appendix.

[11] The quality attribute may capture different characteristics depending on the market. For example in markets for health services quality may be some measure of the length of stay, number of hospital-induced complications, staff per patient, and in education it may reflect the expertise of the teachers or college admission rates. In any case, quality represents a characteristic that cannot be varied on a consumer-by-consumer basis.

[12] It is not necessary that consumers perfectly observe quality as long as the consumers' response to a change in demand is differentiable and predictable by the firm.

[13] Chalkley and Malcomson (1998a) consider the optimal regulatory policy when consumers cannot observe socially valuable quality but the hospital has an altruistic motive, without which the firm would always supply the lowest level of quality.

[14] See Wolinsky (1997), Auriol (1998), Gravelle (1999), and Beitia (2003) for examples of models which specifically utilize structured competition to regulate unverifiable quality.

[15] The assumption that demand is increasing in q means quantity and quality are complements and is not innocuous. If instead quantity and quality are substitutes then the results of the paper must change accordingly.

[16] A capital letter is used for consumer benefit, and later the quality-adjusted consumer benefit to indicate that it is an aggregate measure of all consumers.

[17] Subscripts represent partial derivatives.

line. The cost of production is $\mathcal{C}^2$, increasing, weakly convex in both quantity and quality, and is increasing in θ. Moreover, the marginal cost of output is weakly increasing in quality, $c_{xq} \geq 0$, to reflect the notion that it is more expensive to produce at higher levels of quality. To insure the firm's problem is well-behaved, $c(x, q)$ is strictly quasi-convex. The regulator cannot observe the firm's cost, but knows the functional form.

The firm may also value the degree of community benefit generated by its good, reflecting the mission statement of nonprofit institutions. For example, a hospital may prefer to give up some profit in order to provide the community higher quality health care.[18] The firm's value for community benefit is denoted $\phi(x, q)$ and is concave in its arguments. The firm's value for community benefit can reflect the regulator's valuation of consumer benefit, i.e., $\phi(x, q) = B(x, q)$ or, more likely, it simply reflects the values of the firm's board of directors and is independent of the regulator's valuation for consumer benefit. The firm's objective function is $U(\Pi, \phi) = \beta\Pi + (1-\beta)\phi$, where Π are the firm's profits and $\beta \in [0, 1]$ identifies the relative weight the firm places on profit in relation to community benefit. We further assume that the firm cannot operate with negative profit reflecting the fact that nonprofit institutions cannot operate with consistent losses. This assumption has important ramifications on the regulator's problem when the firm's preference for community benefit is sufficiently strong that the firm bumps up against its non-negative profit constraint.

The Regulator

The regulator is a Stackelberg leader and endowed only with the power to establish a unit price p and transfer payment T. The transfer payment may come from a fixed payment in a two-part tariff and is assumed to not alter the consumers' demand.[19] The prices are enforceable by the regulator.[20]

The regulator's objective is to maximize a weighted average of the expected consumer surplus (CS) and the firm's expected profit (Π). The regulator places a weight $\alpha \in (\frac{1}{2}, 1]$ on consumer surplus and a weight $1 - \alpha$ on the firm's profit.[21]

When the firm possesses superior information about the cost-state, the regulator's uncertainty is represented by the distribution F having strictly positive density f over the support Θ. The characteristics of the regulator's uncertainty are common knowledge.

A. *Quality-Adjusted Cost and Value*

Given a payment mechanism $\{p, T\}$, the firm's objective may be expressed as

$$\max_q \big\{U(p, T; \theta) = \beta\big[px(q, p) - c\big(x(q, p), q; \theta\big) + T\big] + (1 - \beta)\phi\big(x(q, p)\big)\big\} \text{ s.t. } \Pi \geq 0.$$

Let $q(x, p)$ denote the quality-demand function, that is, $q(x, p)$ denotes the level of quality required to induce the equilibrium quantity x given the unit price is set to p. Note that the

[18] Even a nonprofit hospital, which is not permitted to distribute retained earnings to share holders, may earn profit that it uses for managerial perks or other expenditures unrelated to conducting business.

[19] Current Medicare payments only explicitly include the unit payment, though implicitly the lump-sum payment exists through taxes and subsidies.

[20] For example consumers can report to the regulator any instance in which they were charged a different price, or were refused service at the regulated price.

[21] Thus the regulator "cares" more about consumer surplus. Moreover, if $\alpha < 1/2$, then the regulator's problem is maximized with unbounded transfers from consumers to the firm.

properties of $x(q,p)$ are sufficient to insure the existence of $q(x,p)$.[22]

It is intuitive to view the firm's problem as selecting the level of quality that maximizes its profit for a given price and cost state; however, it is equivalent to view the firm's problem as choosing the quantity, x, which maximizes profit given that it must set the quality, $q(x,p)$, in order to induce an equilibrium demand of x. The firm's objective can therefore be alternatively expressed as

$$\max_x \big\{ U(p,T;\theta) = \beta\big[px - c\big(x, q(p,x);\theta\big) + T\big] + (1-\beta)\phi(x) \big\} \text{ s.t. } \Pi \geq 0, \tag{1}$$

where the firm's choice variable is now quantity instead of quality.

Similar to Rogerson (1994),[23] define $g(x;p,\theta)$ as the firm's *quality-adjusted* cost function

$$g(x;p,\theta) = c\big(x, q\big(x,p\big);\theta\big). \tag{2}$$

That is, $g(x;p,\theta)$ denotes the cost of producing the quantity x given that the quality has been adjusted to induce a demand for quantity x when the unit price is p and the cost-state is θ. The relationship between the quality-adjusted marginal cost and the standard marginal cost is

$$\frac{dg}{dx}(x;p,\theta) = \frac{dc}{dx}\big(x, q(x,p);\theta\big) = \frac{\partial c}{\partial x}\big(x, q(x,p);\theta\big) + \frac{\partial c}{\partial q}\big(x, q(x,p);\theta\big)\frac{dq}{dx}(x,p).$$

Thus, the quality-adjusted marginal cost captures both the marginal cost of increasing production, and the marginal cost of increasing the quality necessary to induce the additional demand.

The presence of price in the cost function is unusual, but it allows us to identify a change in cost that occurs as a result of a change in the unit price by the regulator. That is, a change in g with respect to p reflects the change in the firm's cost that follows as a consequence of the firm's reaction to the consumers' price response:

$$\frac{\partial g}{\partial p}(x;p,\theta) = \frac{dc}{dq}\big(x, q(x,p),\theta\big)\frac{dq}{dp}(x,p).$$

It is notable that both g_x and g_p include the term c_q so each only partially captures the change in costs due to a change in the quality level. By using the firm's quality-adjusted cost function we can more clearly identify the change in cost that occurs because the firm chooses to supply more of the service, g_x; and the change in cost that occurs in consequence to a change in the unit price, g_p, which is directly controlled by the regulator.

It should be noted that, given the properties of $c(x,q;\theta)$, $g(x;p,\theta)$ must be $\mathcal{C}^2$, and strictly increasing and convex in x. Moreover, the properties of $c(x,q)$ and $q(x,p)$ further imply $g_{x\theta}(x;p,\theta) \geq 0$; i.e., the marginal cost of output is increasing with the cost state.

Finally, it will also be convenient to define a *quality-adjusted* consumer benefit function V as:

$$V(x;p) = B\big(x, q(x,p)\big). \tag{3}$$

As with the quality adjusted cost function, $V(x;p)$ denotes the consumer benefit to consuming the quantity x given that the quality has been adjusted to induce a demand of x when the unit price

[22] More formally, let $D(x,q,p)$ be the implicit function $x - d(q,p)$, where $d(\cdot)$ has the properties of $x(q,p)$ defined above. Because $d(q,p)$ is continuously differentiable, D has continuous partial derivative D_x, D_p, D_q, and D_θ such that $D_x > 0$ and $D_q > 0$ for all $x > 0$, $q > 0$, $p \geq 0$, and $\theta \in \Theta$. By the Implicit Function Theorem there exists functions f_1 and f_2 such that $x = f_1(q,p)$ and $q = f_2(x,p)$.

[23] In Rogerson (1994) cost is deterministic so $g(\cdot)$ does not take as arguments the price or demand state.

is p. Given the properties of B, it must be the case that V is increasing and concave in x.

II. Social Optimum

We start with the case where the regulator and firm have symmetric information regarding all aspects of the model (e.g., the demand, benefit, and cost functions, the quality, the quantity of output, as well as the cost-state), in order to define the socially optimal outcome. We then proceed to derive the optimal outcome when the regulator and firm have common knowledge of the cost, but the regulator cannot contract directly on output. Although not a pure first-best case, to facilitate the exposition we refer to the solution as first-best when the regulator and firm have symmetric knowledge of the cost state,[24] and we refer to the solution as second-best when the firm has superior knowledge.

The regulator's objective is to maximize the weighted sum of consumer surplus and profit:

$$W(p,T;\theta) = \alpha CS(p,T) + (1-\alpha)\Pi(p,T;\theta). \tag{4}$$

Consumer surplus is defined as the gross consumer surplus minus the cost of the good,

$$CS = V(x;p) - (px + T). \tag{5}$$

By substituting Π and CS into (4) and rearranging, the socially optimal outcome is determined by the maximization program:[25]

$$\max_{x,p,\Pi} \; V\big(x;p\big) - g\big(x;p,\theta\big) - \lambda\Pi$$
$$\text{such that } \Pi \geq 0,$$

where $\lambda = (2\alpha - 1)/\alpha > 0$. By taking the FOCs, the socially optimal outcome is defined as follows.[26]

DEFINITION 1: *The socially optimal outcome consists of the quantity, x^{so}, and prices, $\{p^{so}, T^{so}\}$ that equate both the quality-adjusted social marginal benefit of consumption to the quality-adjusted marginal cost and the quality-adjusted marginal benefit of raising the unit price with the quality-adjusted marginal cost, and additionally leaves the firm with zero profit; that is, prices, output, and profit satisfy:*

$$V_x(x^{so};p^{so}) = g_x(x^{so};p^{so}), \tag{6a}$$
$$V_p(x^{so};p^{so}) = g_p(x^{so};p^{so}), \tag{6b}$$
$$\Pi = 0. \tag{6c}$$

III. Regulating with a Marketed Good

Many regulated markets, including those for health services, require that the consumers pay directly for the good or service.[27] This helps maintain efficiency and avoid over-consumption. When

[24] This is sometimes referred to as a constrained first-best.

[25] A similar optimization program can be derived if the regulator wishes instead to maximize consumer surplus subject to a break-even constraint for the firm.

[26] The properties of V and g imply the objective function is strictly concave.

[27] In markets for health care consumers are generally insured to some degree, but higher medical care costs will invariably lead to higher health insurance premiums so x can be thought of as a crude reduced-form measure of this demand response to

there are no quality considerations, the price fully determines the quantity demanded; however, because the firm is free to adjust the level of quality, it can manipulate demand, reducing the effectiveness of the regulator's pricing rule. This section explores how the firm's ability to manipulate demand affects what the regulator can achieve and establishes the baseline against which we compare the asymmetric information regime.

A. Symmetric Information about θ

When the regulator cannot contract on output, then as a Stackelberg leader, it must offer a contract $\{p, T\}$ which maximizes its objective given how the firm will respond to the payment rule. Denote x^* as the quantity maximizing the firm's objective subject to a nonnegative profit constraint. The first order condition from the firm's problem, (1), shows that the firm's maximizer x^* solves one of two possible equations. First, if the firm's maximization constraint is not binding (i.e., the Lagrange multiplier is zero) we refer to the firm as having a *mixed objective* and x^* solves

$$p + \tfrac{(1-\beta)}{\beta}\phi'\big(x^*(p,\theta)\big) = g_x\big(x^*(p,\theta);p,\theta\big). \tag{7}$$

On the other hand, when the firm's constraint binds then it is a *pure output-maximizer* and x^* solves

$$px^*(p,\theta) + T = g\big(x^*(p,\theta);p,\theta\big). \tag{8}$$

Given the firm's best response function x^*, the regulator's problem (RP-M) can be expressed as[28]

$$\max_{p,\Pi}\; V\big(x^*(p,\theta);p\big) - g\big(x^*(p,\theta);p,\theta\big) - \lambda\Pi \text{ subject to } \Pi \geq 0.$$

The first order condition with respect to p identifies the first-best price p^{fb} as the $p \geq 0$ solving

$$\begin{aligned} V_x\big(x^*(p^{fb},\theta);p^{fb}\big)\tfrac{dx^*}{dp}(p^{fb},\theta) + V_p\big(x^*(p^{fb},\theta);p^{fb}\big) &= \\ g_x\big(x^*(p^{fb},\theta);p^{fb},\theta\big)\tfrac{dx^*}{dp}(p^{fb},\theta) &+ g_p\big(x^*(p^{fb},\theta);p^{fb},\theta\big). \end{aligned} \tag{9}$$

The first-best price consists of the price equating the marginal benefit of increasing the unit price to the marginal cost.

To ease the interpretation of (9), we start by interpreting the marginal cost term. The change in the firm's cost can be thought of as a Slutsky-like decomposition of the change in cost with respect to a price change. The first term on the RHS, $g_x(dx^*/dp)$, identifies the change in the firm's cost that results from adjusting output to take advantage of the change in revenue following

health care costs through the insurance channel.

[28] We have not yet established that the regulator's problem is concave; i.e. $D^2W(p,\Pi) < 0$. To insure concavity in p we must assume the bordered Hessian $|\overline{H}|$ is positive definite. Because the regulator's problem can be expressed equivalently as

$$\max_{x,p,\Pi} V(x;p,\theta) - g(x;p,\theta) \text{ s.t. } \Pi \geq 0 \text{ and } x = x^*(p,\theta).$$

The relevant bordered Hessian (Π enters the regulator's problem linearly so is ignored) is thus defined as

$$|\overline{H}| = \begin{vmatrix} 0 & 1 & -dx^*/dp \\ 1 & V_{xx} - g_{xx} & V_{xp} - g_{xp} \\ -dx^*/dp & V_{xp} - g_{xp} & V_{pp} - g_{pp} + (V_x - g_x)d^2x^*/dp^2 \end{vmatrix} > 0.$$

an increase in price and should be thought of as a *revenue effect*. The second term on the RHS, g_p, identifies the change in the firm's cost associated with raising quality in order to maintain the same quantity of output following a price increase and should be interpreted as a *demand effect* from an increase in price. The marginal benefit consists of the social benefit derived from the firm's adjustment in the equilibrium output, $V_x(dx^*/dp)$, and the benefit gained (or lost) due to the demand effect from a change in unit price.

If a regulatory policy is to achieve the first-best outcome, it must meet two conditions: *(i)*, the firm must be held to zero profits; and *(ii)*, the firm must be induced to produce the efficient quantity x^{fb}. By recognizing that for any p the firm will choose the output equating its marginal benefit of output to its quality-adjusted marginal cost of production the first-best payment policy can be derived. The following proposition reports the optimal payment policy.

PROPOSITION 1: *The optimal payment rule with symmetric cost information for a mixed-objective firm consists of the unique unit price $p^{fb}(\theta)$ and transfer payment $T^{fb}(\theta)$ satisfying:*

$$p^{fb}(\theta) = V_x\big(x^{fb}; p^{fb}\big) + \frac{V_p\big(x^{fb}; p^{fb}(\theta)\big) - g_p\big(x^{fb}; p^{fb}(\theta), \theta\big)}{dx^*\big(p^{fb}(\theta), \theta\big)/dp} - \frac{(1-\beta)}{\beta}\phi'\big(x^{fb}(\theta)\big),$$
$$T^{fb}(\theta) = g\big(x^{fb}; p^{fb}(\theta), \theta\big) - p^{fb}(\theta)x^{fb},$$

and for a pure output-maximizing firm

$$p^{fb}(\theta) = p^{so}(\theta),$$
$$T^{fb}(\theta) = g(x^{fb}; p^{fb}, \theta) - p^{fb}x^{fb},$$

where $p^{fb}(\theta) \leq g_x\big(x^{fb}(\theta); p^{fb}(\theta), \theta\big)$ for all $\theta \in \Theta$.

The payment rules for Proposition 1 simply follows from the first-order condition in the case of the mixed-objectives firm. The restriction that $p^{fb}(\theta) \leq g_x\big(x^{fb}(\theta); p^{fb}(\theta), \theta\big)$ for all $\theta \in \Theta$ follows from lemma 5 in an Appendix.

The intuition for the unit payments are as follows. For a mixed-objective firm, the first two terms for $p^{fb}(\theta)$ account for the revenue and demand effects of adjusting price. Given a unit price, the firm will adjust the quality to insure that its quality-adjusted marginal cost of production equals its marginal benefit; i.e., the unit price. Thus the first term in $p^{fb}(\theta)$ is the marginal social value of changing the equilibrium quantity. The firm adjusts the quality level to compensate for the demand response to the change in price so the second term in p^{fb} is present to account for the change in social value due to the *demand effect* of a change in price. The firm is rewarded for each unit sold and not for the quality level of the good, therefore it must be compensated for each unit of additional surplus and the total change in social surplus due to the demand effect, $(V_p - g_p)$, is divided by the change in the equilibrium quantity, dx^*/dp. The third term adjusts the price to account for the firm's preference for community benefit. The more weight the firm places on community benefit, the less the regulator must compensate the firm to induce the first-best outcome. Lastly, when the firm is a pure output maximizer then the regulator can simply equate the unit price to the socially optimal unit price and adjust the transfer to induce the firm to produce the socially optimal level of output.

It is clear from the definition of $p^{fb}(\theta)$ in Proposition 1, that if the marginal benefit and marginal cost of a price change are equivalent at x^{fb} (i.e., $V_p(x^{fb}; p^{fb}) = g_p(x^{fb}; p^{fb})$), then the first-best and socially optimal unit prices must also be equivalent. Because the firm chooses x

so that $U_x = 0$, that is, because the firm maximizes based on a marginal consumer's valuation of quality, the level of quality will differ from the social optimum. For example, quality will be undersupplied if consumer benefit is the area under the demand curve,[29] or it may be oversupplied if there are negative externalities to consumption.[30] The following proposition formally reports this result.

PROPOSITION 2: *Given a price, p, output may be over- or undersupplied relative to the socially-preferred level. The output differs from the socially preferred output according to the rule:*

$$x^* \gtreqless \arg\max_x (V - g) \text{ when } \quad p \gtreqless V_x(x^*; p) - \tfrac{(1-\beta)}{\beta}\phi'.$$

The reason the regulator is unable to induce the social optimum with symmetric information is because the firm's choice of quality remains non-contractible. In effect, the regulator must use the single instrument of the unit price to control the firm's choice in quality while simultaneously adjusting consumer demand to achieve a socially optimal outcome.

B. Asymmetric Information about θ

Under asymmetric information the problem is a standard adverse selection screening problem, thus to insure there exists a separating equilibrium we require type separation across cost-states. As is common in screening problems, we impose the single crossing property on the firm's value function.

DEFINITION 2: *The single-crossing property holds if the firm's marginal rate of substitution (MRS) of price for transfer payment (U_p/U_T) is monotonic in θ for all $\theta \in \Theta$.*

Without loss of generality, we take advantage of the revelation principle and restrict the analysis to truthful direct mechanisms (Dasgupta et al., 1979; Myerson, 1979). In a direct revelation mechanism the firm announces the state parameter which optimizes its *state-dependent* value function $\mathcal{U}$. Because the firm's objective is to maximize a weighted sum of profit and community benefit, its state-dependent value function is defined as

$$\mathcal{U}(\hat{\theta}, \theta) = \beta\big[p(\hat{\theta})x^*(p(\hat{\theta}), \theta) - g\big(x^*(p(\hat{\theta}), \theta); p(\hat{\theta}), \theta) + T(\hat{\theta})\big] + (1-\beta)\phi\big(x^*(p(\hat{\theta}), \theta)\big),$$

where θ is the true state and $\hat{\theta}$ is the firm's announcement. The regulator's objective is to maximize the total expected social surplus subject to standard individual rationality and incentive compatibility constraints. The regulator's problem may be expressed as

$$\text{(10)} \qquad \max_{p(\theta), \mathcal{U}(\theta)} \int_\Theta \big\{V\big(x^*(p(\theta), \theta); p(\theta)\big) - g\big(x^*(p(\theta), \theta); p(\theta), \theta\big) - \tfrac{\lambda}{\beta}\big(\mathcal{U}(\theta) - (1-\beta)\varphi(\theta)\big)\big\} dF(\theta),$$

subject to

[29] This is the analog to a result first reported by Spence (1975) and Baron (1981), who studied the provision of quality in a more general framework.

[30] In Baron (1981) the regulator's value function is simply the area under the demand curve; i.e., $B = \int_0^x P(\tilde{x}, q) d\tilde{x}$ where P is the inverse demand function. Thus, without externalities, $V_x = p + \int_0^x P_q(\tilde{x}, q) q_x \, d\tilde{x} \neq p$ and $V_x > p^{fb} = g_x$ implying both output and quality are undersupplied. If the regulator's measure of consumer benefit accounts for some negative externality then it may be the case that $V_x < p$. As an example consider the social surplus function characterized as a downward, parallel shift of the demand curve: $S = \int_0^x (P(\tilde{x}, q) - \beta) d\tilde{x}$. The FOC yields, $V_x = p - \beta + \int_0^x P_q(\tilde{x}, q) q_x \, d\tilde{x}$. Thus, $V_x < p$ if $\beta > \int_0^x P_q(\tilde{x}, q) q_x \, d\tilde{x}$; that is, the quality adjusted marginal surplus is less than the unit price if the cost of the negative externality exceeds the marginal benefit of the change in consumption caused by a change in the quality.

$$\begin{array}{lll} \mathcal{U}(\theta) \geq 0 & \forall\theta \in \Theta & \text{(Individual Rationality)} \\ \mathcal{U}(\theta,\theta) \geq \mathcal{U}(\hat{\theta},\theta) & \forall\hat{\theta},\theta \in \Theta & \text{(Incentive Compatibility).} \end{array}$$

where $\mathcal{U}(\theta) = \mathcal{U}(\theta,\theta)$ and $\varphi(\theta) = \phi\big(x^*\big(p(\theta),\theta\big)\big)$. Note that the firm's rents are expressed as the difference between the firm's total value for output, $\mathcal{U}(\theta)$, and the firm's value for community benefit given a truthful report of the cost state, $\varphi(\theta)$. The expression $\mathcal{U}(\theta) - (1-\beta)\varphi(\theta)$ is divided by β to account for the fact that the firm places weight β on profit whereas the regulator is concerned about total social loss from profit.

The following lemma characterizes the necessary and sufficient conditions for an incentive compatible payment policy.

LEMMA 1: *The menu of two-part tariffs $\{p(\theta), T(\theta)\}_{\theta\in\Theta}$ is incentive compatible if and only if it satisfies the conditions*

$$\textit{(i)}\ \ \frac{d\mathcal{U}(\theta)}{d\theta} = \frac{\partial\mathcal{U}(\theta)}{\partial\theta} = \begin{cases} -\beta g_\theta\big(x^*(p(\theta),\theta);p(\theta),\theta\big) & \textit{when } \Pi(x^*) > 0, \\ -(1-\beta)\phi'(x)\dfrac{g_\theta(x^*;p,\theta)}{g_x(x^*;p,\theta)-p} & \textit{otherwise.} \end{cases}$$

$$\textit{(ii)}\ \ \textit{sign}[dp\,/d\theta] = \textit{sign}\left[\tfrac{\partial}{\partial\theta}\left(U_p/U_T\right)\right].$$

Condition *(i)* is critical to the results and is a consequence of the fact the firm will lower the service quality in higher cost-states unless it receives a higher unit payment to compensate for higher marginal costs. Providing a higher unit payment in higher states generates an incentive for the firm to misreport the state as higher than it is. Consequently incentive compatibility requires the firm receive higher rents in lower cost states to counter the desire to misreport the cost state as being high. When the firm's optimal choice leaves it with some positive profit, then anything that increases the rate at which the firm's costs increase with the state will cause it to receive higher rents in all states. When the firm does not earn positive profit then it chooses the output that leaves it with zero profit. Again, the firm has an incentive to misreport the state if doing so allows it to increase output. The fraction $g_\theta/(g_x - p)$ identifies the trajectory for the firm's choice of output across cost states, which notably is a direct function of the unit price. Both conditions are negative, thus the firm's informational advantage is necessarily decreasing with the cost state.

Condition *(ii)* is specific to the use of the two-part tariff and reports that any payment policy satisfying incentive compatibility must offer a unit payment that moves in the same direction as the firm's MRS of price for the fixed transfer. Because the firm's MRS of price for fixed transfer must be monotone for type separation, it follows that the regulator's pricing rule is also monotone; however, it is not restricted to being either an increasing or a decreasing function of the state parameter.

Using Lemma 1 we can also identify the firm's expected information rents when its optimal choice of output leaves it with positive profit. The following corollary reports this result.

COROLLARY 1: *An incentive compatible menu of two-part tariffs $\{p(\theta), T(\theta)\}_{\theta\in\Theta}$ must leave a marginal optimizer with expected information rents:*

$$E\Pi(\theta) = \int_{\underline{\theta}}^{\overline{\theta}} \big\{ \tfrac{1-\beta}{\beta}\big[\varphi(\overline{\theta}) - \phi\big(x^*(p(\theta),\theta)\big)\big] + \tfrac{F(\theta)}{f(\theta)} g_\theta(x^*;p,\theta)\big\} dF(\theta). \tag{11}$$

When the firm's optimal choice of output leaves it with some positive profit, the firm acts in a similar manner to a pure profit-maximizing firm in that it increases output until the marginal benefit of increasing output equals the marginal cost. The more weight the firm places on community benefit, the more it helps the regulator by limiting its own rents.

If the firm's preference for community benefit causes the non-negative profit constraint to bind, then the regulator cannot focus on the firm's rents in any cost state, but must instead insure that the firm does not misreport the cost state in order to over-produce for the sake of output. As a consequence, the approach to solving the optimal contract differs depending on if the firm has a mixed objective, or is purely output-maximizing. The next two subsections analyzes the optimal contract for a firm with mixed objectives, and a purely output-maximizing firm, respectively.

A Mixed-Objectives Firm

We will solve the regulator's problem using optimal control. The regulator's problem is to maximize Eq. (10) subject to the constraints

$$\begin{aligned} d\mathcal{U}/d\theta &= -\beta g_\theta\big(x^*(p(\theta),\theta);p(\theta),\theta\big), \\ \mathcal{U}(\overline{\theta}) &= \varphi(\overline{\theta}). \end{aligned}$$

The Hamiltonian for the regulator's problem can thus be expressed as

$$\begin{aligned} H = \Big\{ V\big(x^*(p(\theta),\theta);p(\theta)\big) - g\big(x^*(p(\theta),\theta);p(\theta),\theta\big) - \tfrac{\lambda}{\beta}\big(\mathcal{U}(\theta) - (1-\beta)\phi\big(x^*(p(\theta),\theta)\big)\big)\Big\} f(\theta) \\ - \delta(\theta)\beta g_\theta\big(x^*(p(\theta),\theta);p(\theta),\theta\big), \end{aligned}$$

where $\mathcal{U}(\theta)$ is the state variable, $p(\theta)$ is the control, and $\delta(\theta)$ is the Pontryagin multiplier. By the maximum principle $dH/d\mathcal{U} = -d\delta/d\theta = -(\lambda/\beta)f(\theta)$. The boundary condition at $\underline{\theta}$ is unconstrained so the transversality condition at $\theta = \underline{\theta}$ is $\delta(\underline{\theta}) = 0$. Integrating $d\delta/d\theta$ yields $\delta(\theta) = (\lambda/\beta)F(\theta)$.

Plugging in $\delta(\theta)$ and taking the first-order condition of the Hamiltonian yields

$$\begin{aligned} \tfrac{d}{dp}\big\{V\big(x^*(p(\theta),\theta);p(\theta)\big) - g\big(x^*(p(\theta),\theta);p(\theta),\theta\big)\big\} = \\ \lambda\left[\tfrac{F(\theta)}{f(\theta)}\tfrac{d}{dp}\{g_\theta(x^*;p,\theta)\} - \tfrac{1-\beta}{\beta}\phi'(x^*)\tfrac{dx^*}{dp}\right]. \end{aligned} \tag{12}$$

The price p^{sb} solving Eq. (12) is the second-best price given the regulator's constraints and (12) equates the marginal change in social surplus to the marginal social loss. Eq. (12) has a familiar interpretation. Increasing the payment to the firm in cost-states $[\theta, \theta + d\theta]$, which number $f(\theta)d\theta$, by dp increases the social surplus by $\big[V_p(x^*;p,\theta) - g_p(x^*;p,\theta) + \big(V_x(x^*;p,\theta) - g_x(x^*;p,\theta)\big)\frac{dx^*}{dp}\big]dp$. From *(i)* of Lemma 1, the increase in output simultaneously increases the firm's rent in cost-states $[\underline{\theta}, \theta]$, which number $F(\theta)$, by $\frac{d}{dp}\{g_\theta(x^*;p,\theta)\}dp$, offset by the degree to which the firm's value for community benefit changes with respect to the increase in price, $\frac{1-\beta}{\beta}\phi'(x^*)\frac{dx^*}{dp}f(\theta)dp$. The total social cost of the increase in the firm's rent is $\lambda\big\{F(\theta)\frac{d}{dp}\{g_\theta(x^*;p,\theta)\} - \frac{1-\beta}{\beta}\phi'(x^*)\frac{dx^*}{dp}f(\theta)\big\}dp$.

If the price p^{sb} solving Eq. (12) does not satisfy condition *(ii)* of Lemma 1 then the payment rule is not incentive compatible and the regulator is unable to extract any usable information from the firm. Consequently, permitting the firm to use its private knowledge to select the unit price is too socially costly. In such a case the regulator cannot allow the firm to choose the price and must set a fixed-price, $p(\theta) = \hat{p} \in [p^{sb}(\underline{\theta}), p^{sb}(\overline{\theta})]$.[31] Caillaud et al. (1988) and Laffont and Tirole (1993,

[31] This represents an analog to the case of decreasing marginal costs in Lewis and Sappington (1988) and the case of an

pg. 161) refer to this as "the phenomenon of nonresponsiveness of the allocation with respect to private information."

Identifying if condition *(ii)* of Lemma 1 is satisfied requires either establishing functional forms for the cost, value, and demand functions or making several strict assumptions on the higher order derivatives of these functions.[32,33] To maintain generality and avoid making assumptions on higher-order derivatives that do not have an economic justification, for the remainder of this section we will assume the properties of the cost, value, and demand functions are sufficient to insure that the relationship reported in Lemma 1 is satisfied. See the Appendix for more discussion on when the payment rule will be incentive compatible and allow for full type separation.

Returning to the first-order condition of the regulator's problem given by Eq. (12) we can return to the focus of this study and analyze how the firm's informational advantage distort output. Recall that the first-best solution is the contract that sets the left-hand side of (12) equal to zero. Because the firm's information rent may be either increasing or decreasing with the unit payment, however, the right-hand side of (12) may be either less than or greater than zero, resulting in either $p^{sb} > p^{fb}$ or $p^{sb} < p^{fb}$, respectively. Intuitively, the regulator will shade the price up or down from the first-best price to limit the information rents attained by the firm. That is, if the point-wise derivative of the firm's rents with respect to price is increasing with the unit payment ($dE\Pi/dp > 0$), then the regulator will decrease the payment to limit the firm's rent, and vice-versa. The following proposition formally reports this relationship between first-and second-best unit payments.

PROPOSITION 3: *The relative magnitude of the second- to the first-best unit price is inversely related to the effect a price change has on the firm's information rent:*

$$p^{sb}\begin{cases} < p^{fb} & \textit{when} \quad \frac{F(\theta)}{f(\theta)} d\{g_\theta(x^*;p,\theta)\}/dp > \frac{1-\beta}{\beta}\phi'(x^*)\frac{dx^*}{dp}, \\ = p^{fb} & \textit{when} \quad \frac{F(\theta)}{f(\theta)} d\{g_\theta(x^*;p,\theta)\}/dp = \frac{1-\beta}{\beta}\phi'(x^*)\frac{dx^*}{dp}, \\ > p^{fb} & \textit{when} \quad \frac{F(\theta)}{f(\theta)} d\{g_\theta(x^*;p,\theta)\}/dp < \frac{1-\beta}{\beta}\phi'(x^*)\frac{dx^*}{dp}. \end{cases}$$

Notably the conditions reported by proposition 3 indicate that the second-best price may be distorted away from first-best for *all* types. This is in contrast to the situation with a pure profit-maximizing firm. When the firm is a pure profit-maximizer, the classic result is obtained and there will be no distortion in the price in the lowest cost state $\underline{\theta}$.

More importantly, it follows that if the second-best unit price may be higher or lower than the first-best, then the second-best output may be under- or over-supplied relative to the first-best as well. For example, when the second-best price is less than the first-best, and the firm's optimal choice of output is increasing in the price ($dx^*/dp > 0$), then the second-best level of output will necessarily be below first-best and when the firm's optimal choice of output is decreasing in the price ($dx^*/dp < 0$), then the second-best level of output will exceed the first-best output. When the second-best price is greater than the first-best the results are of course reversed.

It comes as no surprise that by adding $\phi(x)$ as a general expression of the firm's valuation of community benefit we can achieve an ambiguous result. For example, if a hospital's board of directors values community benefit substantially more than the regulator then the regulator may have no choice but to allow the hospital to over-supply hospital services to some degree in order

increasing labor allocation in a self-managed firm in Guesnerie and Laffont (1984). In both cases the optimal regulatory policy fails incentive compatibility eliminating the regulator's ability to extract any information about the state of the world.

[32]For example,the signs for $g_{\theta\theta p}$, $g_{\theta px}$, $g_{\theta xx}$ and $g_{\theta\theta x}$ must be established.

[33]See Rogerson (1987) for a discussion on the necessary properties of the model's primitives that allow for an implementable payment policy for a similar principal-agent problem.

to maintain incentive compatibility. It is surprising to note, however, that this outcome is not necessarily dependent on the characteristics of ϕ; that is, even with a pure profit-maximizing firm, the second-best outcome may be under- *or* over-supplied relative to first-best. The direction of the distortion is determined by how the firm's choice of output and rents are altered by a change in the unit payment. To see this, we start with the definition of the firm's ex ante information rents identified in corollary 1. When the firm is a pure profit-maximizer, then $\beta = 1$ and taking the point-wise derivative of the firm's expected rents yields:

$$\frac{d}{dp}\left\{\frac{F(\theta)}{f(\theta)}\frac{\partial g(x^*;p,\theta)}{\partial\theta}\right\} = \frac{F(\theta)}{f(\theta)}\left[\frac{\partial^2 g}{\partial\theta\partial p} + \frac{\partial^2 g}{\partial\theta\partial x}\frac{dx^*}{dp}\right].$$

The first term within the brackets on the RHS, $g_{\theta p}$, identifies the direct change to the firm's rent that follows from an increase in the unit price. The partial change to the firm's rents following a price change (in all cost states) is the change to the firm's rent that comes about from adjusting quality to maintain the *same* equilibrium output. Increasing quality increases the cost of production, which in turn increases the incentive to misreport a high cost-state. Therefore the firm's rents must increase in all states in the price dimension, $g_{\theta p} \geq 0$. The second term $g_{\theta x}(dx^*/dp)$ identifies the indirect change in the firm's rents that follows from adjusting the equilibrium quantity demanded due to a price change. Increasing output increases the cost of production and the firm's rents increase in all states along the quantity dimension as well. Therefore, $d\{g_\theta\}/dp < 0$ if and only if the firm's best response to a price *increase* is to *decrease* the equilibrium quantity sufficiently as to lower its overall costs in every state θ. That is, the firm's rents are decreasing with a price increase if and only if $dx^*/dp < -(g_{\theta p}/g_{\theta x}) < 0$.

When $dx^*/dp > 0$ the second-best output is undersupplied relative to first-best because the firm's rents are unambiguously increasing with the unit payment. The regulator thus sets a price below the first-best price in order to limit the firm's rents, resulting in the undersupply. On the other hand, when $dx^*/dp < 0$ the outcome depends on whether or not dx^*/dp is sufficiently negative to flip the firm's rents so that they are decreasing in the unit payment. If it is, then the regulator will have to set a price above the first-best price to limit the firm's rents. Because the firm's optimal choice of output varies inversely with the price, the higher unit payment causes the firm to still choose an output below the first-best. Furthermore, if dx^*/dp is insufficiently negative, then the firm's rents are still increasing with the price and the regulator will again choose a price below first-best. Because the firm's best response to a decrease in price is to *increase* output, this results in an oversupply relative to first-best. The following proposition formally identifies these three cases.[34]

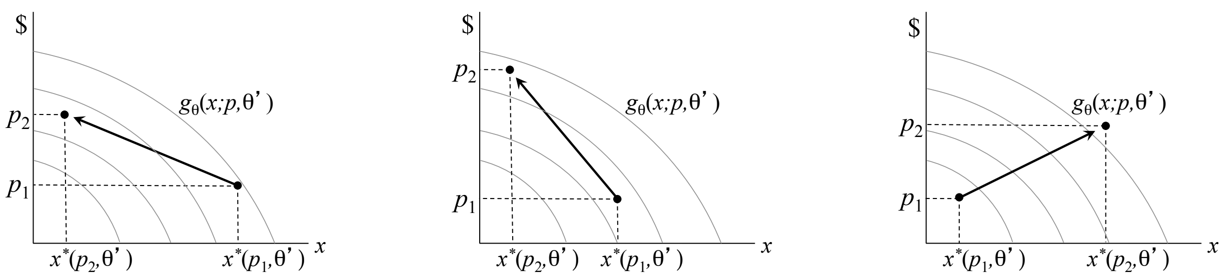
PROPOSITION 4: *With asymmetric cost information and a pure profit-maximizing firm, the relative size of the second- to first-best price and output is determined by the rules:*

$$\begin{array}{rllll}
(i) & dx^*/dp < -(g_{\theta p}/g_{\theta x}) < 0 & \Rightarrow d\{g_\theta\}/dp < 0 \Rightarrow p^{sb}(\theta) > p^{fb}(\theta) & \Rightarrow x^{sb}(\theta) < x^{fb}(\theta), \\
(ii) & -(g_{\theta p}/g_{\theta x}) < dx^*/dp < 0 & \Rightarrow d\{g_\theta\}/dp > 0 \Rightarrow p^{sb}(\theta) < p^{fb}(\theta) & \Rightarrow x^{sb}(\theta) > x^{fb}(\theta), \\
(iii) & -(g_{\theta p}/g_{\theta x}) < 0 < dx^*/dp & \Rightarrow d\{g_\theta\}/dp > 0 \Rightarrow p^{sb}(\theta) < p^{fb}(\theta) & \Rightarrow x^{sb}(\theta) < x^{fb}(\theta),
\end{array}$$

for any $\theta \in (\underline{\theta}, \overline{\theta}]$ *and at* $\underline{\theta}$, $p^{sb}(\underline{\theta}) = p^{fb}(\underline{\theta})$ *and* $x^{sb}(\underline{\theta}) = x^{fb}(\underline{\theta})$.

[34] Identification of the output distortion requires one more technical assumption that $sign[dx^*(p^{sb})/dp] = sign[dx^*(p^{fb})/dp]$; i.e., the price distortions are not dramatic enough that the firm's output response to a price change moves in opposite directions at the first- and second-best prices. Though the qualitative results will remain, relaxing this assumption introduces additional cases.

Figure 1 graphically represents the three cases identified by Proposition 4. The horizontal axis represents the quantity of output, and the vertical axis the price in dollars. The level curves for a representative $g_\theta(x^*; p, \theta')$ are displayed and should be thought of as iso-rent curves since the firm's information rents are a function of how its costs change with the state parameter. The iso-rent curves are increasing away from the origin; i.e., $g_{\theta p} > 0$ and $g_{\theta x} > 0$. In Figures 1(a) and 1(b) the firm's optimizer decreases with an increase in price and in figure 1(c) it increases. Figure 1(a) represents case *(i)* of proposition 4 as the decrease in output is less than the MRS, thus sufficient to drop to a lower level curve indicating the firm's information rent decreases. In Figure 1(b) the decrease in output is insufficient. Representing case *(ii)*, the decrease in output is higher than the MRS resulting in a jump to a higher level curve. Figure 1(c) corresponds with case *(iii)* as the change in output following a price increase also results in a straightforward jump to a higher level curve indicating an increase in the firm's rents. It should be noted that movement across iso-rent curves need not be monotone across cost states when the firm's optimal choice of output is decreasing with price. Thus, for some functional forms, the regulator's problem will satisfy condition *(i)* in some states and for others condition *(ii)* of Proposition 4.



(a) Case *(i)* (b) Case *(ii)* (c) Case *(iii)*

FIGURE 1. GRAPHICAL EXAMPLES OF THE PRICING RULES FROM PROPOSITION 4

Importantly proposition 4 informs us that we cannot predict *a priori* how the firm's informational advantage will distort output without having functional forms. This qualitative result is in sharp contrast to the results found in the variety of research cited in the introduction where the firm's informational advantage results in a strictly downward distortion from first-best. With its choice of quality, the firm is able to adjust the quantity demanded for a given unit price. In consequence, the unit price does not just alter the quantity demanded by consumers, but it alters the quality level the firm chooses, which in turn also alters the quantity demanded. The two sources of demand adjustment may lead to either an over- or under-supply relative to first best. Identifying the distortion to quality is more problematic. For example, in case *(iii)* it is clear that if prices and output are below first-best levels, then quality must be undersupplied, but in cases *(i)* and *(ii)* quality may be under- or oversupplied relative to first-best. When the information asymmetry is in demand, however, we show in Proposition 10 (in an Appendix) that quality may be unambiguously oversupplied relative to first-best.

We will again use optimal control to analyze the regulator's problem; however, we must choose a different state variable because the firm solves half of the regulator's problem by leaving itself with zero profit and the regulator does not care about the firm's utility from output.

Because the regulator cares about maximizing social surplus let $S(\theta) \equiv V(x;p,\theta) - g(x;p,\theta)$, be the state variable and let $p(\theta) = p^{np}(\theta)$ be the control variable. The regulator's objective is to maximize $\int_\Theta S(\theta)dF(\theta)$ subject to

$$dS\,/d\theta = \tfrac{d}{d\theta}\big\{V - g\big\} = \big(V_x - g_x\big)\tfrac{dx^*}{d\theta} + \big(V_p - g_p + (V_x - g_x)\tfrac{dx^*}{dp}\big)\tfrac{dp^{np}}{d\theta} - g_\theta. \tag{13}$$

Let $\delta(\theta)$ be the Pontryagin multiplier, then by the maximum principle $dH\,/d\mathcal{S} = -d\delta\,/d\theta = -f(\theta)$. The boundary condition at $\underline{\theta}$ is unconstrained so the transversality condition at $\theta = \underline{\theta}$ is $\delta(\underline{\theta}) = 0$. Thus, integrating $d\delta\,/d\theta$ yields $\delta(\theta) = F(\theta)$. The first-order condition of the Hamiltonian yields:

$$\tfrac{d}{dp}\big\{\big(V_x - g_x\big)\tfrac{dx^*}{d\theta} + \big(V_p - g_p + (V_x - g_x)\tfrac{dx^*}{dp}\big)\tfrac{dp^{np}}{d\theta} - g_\theta\big\} = 0, \tag{14}$$

where p^{np} is the unit price solving the first-order condition. When $dx^*\,/dp = dx^{fb}\,/dp$ and $dp^{np}\,/d\theta = dp^{fb}\,/d\theta$ then the expression on the LHS of (14) is, by definition, equal to zero. Without using a restricted functional form for the cost and value functions there does not exist a closed form solution to the regulator's problem.[35] It can be shown, however, that the regulator will not generically be able to achieve the first-best with an output-maximizing firm. Recall that Lemma 1 indicates that $dx^*\,/d\theta = -g_\theta(x^*;p,\theta)/\big(g_x(x^*;p,\theta) - p\big)$. Thus the trajectory of x^* across cost-states is a direct function of the unit payment. Lemma 5 (in an Appendix) reports that $p \leq g_x$ for an output maximizing firm, therefore $dx^*\,/d\theta < 0$ and, using the integrable form of the envelope theorem (Milgrom, 2001, pg. 67), $x^*(\theta) = x^*(\hat{\theta}(\theta),\theta)$ may be expressed as

$$x^*\big(\hat{\theta}(\theta),\theta\big) = x^*\big(\hat{\theta}(\overline{\theta}),\overline{\theta}\big) - \int_\theta^{\overline{\theta}} \frac{\partial x^*}{\partial\theta}\big(\hat{\theta}(\tilde{\theta}),\tilde{\theta}\big)d\tilde{\theta},$$

where $\partial x^*/\partial\theta = -g_\theta(g_x - p)$ as established by Lemma 1.

Thus, if the regulator can design a contract so that $x^*(\overline{\theta}) = x^{fb}(\overline{\theta})$ and $dx^*\,/d\theta = dx^{fb}\,/d\theta$ in all $\theta \in \Theta$ then it will induce the first-best output in all θ. The following proposition identifies the unique payment policy and accompanying necessary and sufficient conditions that accomplishes this outcome.[36]

PROPOSITION 5: *The menu of two-part tariffs $\{p^{np}(\theta), T^{np}(\theta)\}_{\theta\in\Theta}$ induces an output-maximizing firm to produce the first-best quantity x^{fb}, where*

$$p^{np}(\theta) = g_x\big(x^{fb}(\theta);p^{np}(\theta),\theta\big) + \frac{g_\theta\big(x^{fb}(\theta);p^{np}(\theta),\theta\big)}{dx^{fb}(\theta)\,/d\theta},$$
$$T^{np}(\theta) = g\big(x^{fb}(\theta);p^{np}(\theta),\theta\big) - p^{np}(\theta)x^{fb}(\theta),$$

[35] For this reason we cannot identify how the second-best price is distorted away from the first-best price beyond demonstrating that they are not equal.

[36] Existence of a p^{np} that solves the payment rule in Proposition 5 is not guaranteed for all quality adjusted cost functions, g. For example, if $g_\theta/(dx^{fb}\,/d\theta)$ is too negative then there may not exist a $p^{np}(\theta) > 0$ that induces the first-best output in all θ when the good is marketed. There always exists a $p^{np}(\theta)$ when the good is nonmarketed, however, because in principle $p^{np}(\theta)$ can be negative.

if and only if $dx^{fb}/d\theta < 0$, $dp^{np}/d\theta < 0$, *and* $p^{np}(\theta) \geq 0$ *for all* $\theta \in \Theta$.

The payment rule reported by Proposition 5 is sufficient to induce the first-best *output*; however, it does not induce the first-best *outcome*. This follows because the equilibrium quantity is determined by both the price consumers pay and the level of quality established by the firm. Because the regulator is constrained to use the unit payment to control the firm's choice of output, x^*, it cannot also set the price to the first-best price. Given the concavity of the regulator's problem the first-best unit price is unique as is the payment policy inducing the first-best output, and they are not generically equivalent. Put another way, the firm has one instrument, with which to solve two equations; and it cannot. The following proposition formally states this result.

PROPOSITION 6: *The regulator cannot generically induce the first-best outcome for a pure output-maximizing firm.*

Note that because the payment policy identified by Proposition 5 does not induce the first-best outcome with a marketed good, it is must be the case that the payment rule that induces $dx^{fb}/d\theta$ does not solve (14) and is also not the second-best payment rule.

IV. A Nonmarketed Good

When the good or service is not marketed then there is no direct demand-response to the contract and the regulator only needs to account for how the firm best-responds to the payment rule it sets.

A. Symmetric Information about θ

The regulator's problem continues to be that of designing a menu of two-part tariffs which maximizes a weighted sum of consumer surplus and profit given the firm will choose the quantity solving (8). Because payment is made directly by the regulator using public funds raised through taxation, we introduce a shadow cost to public funds, $\gamma > 0$. In this way, every \$1 paid by the regulator for the firm's good has a total social cost of $\$(1+\gamma)$. Net consumer surplus is now defined as $CS = B(x,q) - (1+\gamma)(px+T)$.

As with a marketed good, it is convenient to define and work with the quality-adjusted cost and consumer value functions. Consumer demand is unaffected by the unit price, though, so the quality-adjusted functions no longer take price as an argument; i.e., $g(x;\theta) = c(x, q(x);\theta)$ and $V(x) = B(x, q(x))$.

The regulator's problem (RP-NM) with symmetric cost information can be expressed as

$$\max_{p,\Pi} \; V\big(x^*(p,\theta)\big) - (1+\gamma)g\big(x^*(p,\theta);\theta\big) - (\lambda+\gamma)\Pi \quad \text{subject to } \Pi \geq 0.$$

The firm's profit still enters the regulator's problem negatively; however, the shadow-cost to public funds increases the loss to social welfare that positive firm profit generates. Nevertheless, removing the unit price from consumer demand simplifies the regulator's problem. For example, because the objective function is strictly concave in x, it is strictly concave in p without any further assumptions.[37] Moreover, the firm's best response to a price increase is to increase output.[38] This follows

[37]To see why, start with $d^2\{V-g\}/dp^2 = (V_{xx} - g_{xx})\frac{dx^*}{dp}$. The term in parenthesis is strictly negative thus concavity only requires $dx^*/dp > 0$. Whereas dx^*/dp could be either positive or negative when demand is a function of price, it is strictly positive for a nonmarketed good.

[38]From the conjugate pairs theorem $sign[dx^*/dp] = sign[\Pi_{xp}]$ and $\Pi_{xp} = 1$.

because an increase in the unit price increases the firm's revenue with no concomitant increase in cost, therefore the firm will increase output until its marginal cost of production again equals the higher unit payment.

The first order condition of the regulator's problem (RP-NM) yields the first-best payment rule with symmetric information.

PROPOSITION 7: *The optimal payment rule for a nonmarketed good with symmetric cost and demand information for a mixed-objectives firm consists of the unique unit price $p_{nm}^{fb}(\theta)$ and transfer payment $T_{nm}^{fb}(\theta)$ satisfying*

$$\begin{aligned} p_{nm}^{fb}(\theta) &= \tfrac{1}{1+\gamma} V_x\big(x^{fb}\big) - \tfrac{(1-\beta)}{\beta}\phi'\big(x^{fb}(\theta)\big), \\ T_{nm}^{fb}(\theta) &= g\big(x^{fb}, \theta\big) - p_{nm}^{fb}(\theta) x^{fb}, \end{aligned}$$

and for a pure output-maximizing firm

$$\{p_{nm}^{fb}(\theta), T_{nm}^{fb}(\theta)\} \in \big\{\{p, T\} \mid 0 \leq p \leq g_x(x^{so}, \theta) \text{ and } T = g(x^{so}, \theta) - p x^{so}\big\},$$

for all $\theta \in \Theta$.

Because price is not present in the demand function, the first-best solution simply equates the marginal benefit of additional consumption with the social marginal cost. More importantly, because the socially optimal level of output is determined only by the firm's service quality, the first-best and socially optimal outcomes are equivalent and by simply modifying the way consumers pay for the service, the regulator may be able to improve the outcome.[39] Notably, when the firm is a pure output-maximizer, the optimal contract is no longer unique, suggesting the regulator will have additional flexibility to induce its preferred outcome when the firm has superior information.

B. Asymmetric Information about θ

When information is asymmetric, the regulator's problem is to maximize

$$\max_{p(\theta), U(\theta)} \int_{\Theta} \big\{ V\big(x^*(p(\theta), \theta), \theta\big) - (1+\gamma) g\big(x^*(p(\theta), \theta), \theta\big) - \tfrac{(\lambda+\gamma)}{\beta}\big(U(\theta) - (1-\beta)\varphi(\theta)\big)\big\} dF(\theta),$$

subject to individual rationality and incentive compatibility constraints similar to those for the marketed good.

The lack of a demand response to price also simplifies the regulator's problem under asymmetric information because the SCP is automatically satisfied as reported by Lemma 2.

LEMMA 2: *When the good is nonmarketed, the SCP is satisfied and $d\{\mathcal{U}_p/\mathcal{U}_T\}/d\theta < 0$ for all $\theta \in \Theta$.*

Any incentive compatible mechanism must still satisfy Lemma 1.[40] Because the MRS of the unit payment for the fixed transfer is increasing in the cost parameter, another simplification to the regulator's problem due to the lack of price response is that condition *(ii)* of Lemma 1 reduces to the following.

[39] The observation that a nonmarketed good does not lead to an outcome distorted away from the social optimum was first made by Ma (1994) and Chalkley and Malcomson (1998b) using similar models.

[40] Note that the proof for Lemma 1 holds when $\partial g/\partial p = 0$.

LEMMA 3: *When the good is nonmarketed, the payment policy* $\{p(\theta), T(\theta)\}_{\theta\in\Theta}$ *is incentive compatible only if* $dp/d\theta \leq 0$*.*

Intuitively, the regulator must set a lower unit payment in higher cost states and compensate the firm via more of the fixed transfer to remove any incentive to misreport the cost as being higher than it is.

A Mixed-Objectives Firm

The regulator's problem can again be solved utilizing optimal control. The first-order condition of the Hamiltonian yields

$$V_x\big(x^*\big(p(\theta),\theta\big),\theta\big) - (1+\gamma)g_x\big(x^*\big(p(\theta),\theta\big),\theta\big) = (\lambda+\gamma)\left[\frac{F(\theta)}{f(\theta)}g_{\theta x}(x^*;\theta) - \frac{1-\beta}{\beta}\phi'(x^*)\frac{dx^*}{dp}\right]. \tag{15}$$

The quantity x^* solving Eq. (15) is the second-best quantity given the regulator's constraints. The interpretation of Eq. (15) is similar to the interpretation of Eq. (12), the first-order condition for a marketed good. When the firm has a mixed objective the direction of the distortion again depends on whether the firm's information rents are increasing or decreasing with the unit payment and Proposition 3 continues to apply; however, the lack of a demand response to price simplifies the results in two ways. First, because $dx^*/dp > 0$ the output will always be distorted in the same direction as the unit payment as defined in Proposition 3. Second, because $d\{g_\theta(x^*;\theta)\}/dp$ $(= g_{\theta x}(dx^*/dp))$ is unambiguously positive, the second-best unit price and equilibrium quantity are distorted *strictly* downward from the first-best levels with a pure profit-maximizing firm. This result is stated in the following proposition.

PROPOSITION 8: *When the good is nonmarketed and the firm is profit-maximizing, the second-best unit price and equilibrium quantity are distorted downward from the first-best for all but the lowest state,* $\underline{\theta}$ *where they are equivalent.*

The proof follows immediately from Proposition 3. The removal of a demand response reduces the characteristics of the problem to rule *(iii)* of Proposition 4 and can be graphically represented by figure 1(c) by transforming the level curves into vertical lines.

Output-Maximizing Firm

Lastly, when the firm's preference for community benefit is so strong that it is a pure output-maximizer, then the payment rule established by Proposition 5 is still the unique payment rule inducing the first-best output when the good is nonmarketed. Moreover, as there is no demand response to the unit price, the equilibrium output is uniquely determined by quality. Thus, as long as the unit payment is decreasing with the cost-state ($dp^{np}/d\theta < 0$) for incentive compatibility, the regulator is free to use the unit price to induce the firm to produce in the social interest and the payment rule reported by Proposition 5 induces the first-best outcome. This is stated in the following proposition.

PROPOSITION 9: *When the good is nonmarketed and the firm is output-maximizing the regulator can induce the first-best outcome using the payment rule reported by Proposition 5 with the exception that* $p(\theta)$ *may take a negative value.*

In addition to the inelasticity of demand to the unit payment in the non-marketed case, another critical difference between the marketed and non-marketed cases is the possibility of the unit payment taking a negative value. When the good or service is non-marketed this serves to increase the marginal cost of output and is necessary only when the first-best level of output decreases in a sufficiently strong manner across cost states; i.e., when $dx^{fb}/d\theta < -\big(g_\theta(x^{fb};\theta)/g_x(x^{fb};\theta)\big)$.

V. Concluding Remarks

This paper has examined the optimal payment policies for a monopolist who can manipulate demand through its choice of unverifiable quality. We have assumed that the regulator cannot contract on quality, output, or the firm's cost ruling out many contracting regimes such as rate-of-return or minimum quality standards regulation. We have found that within the same informational environment the regulator can achieve strikingly different outcomes based on the consumers' access to the good and the firm's objective. Indeed, when the good is nonmarketed and the firm is a pure output-maximizer the regulator can completely attenuate the informational advantage of the firm. Because there is a deadweight loss associated with using public funds, however, eliminating the firm's informational advantage includes a social cost and does not represent a panacea for the regulator. In contrast, when the good is marketed the equilibrium output may be under- or oversupplied relative to first-best—even for a pure profit-maximizing firm.[41]

The ambiguous direction of the distortion that occurs with a marketed good follows directly from the fact that the firm can manipulate demand through its choice of quality. If the regulator raises the unit payment, then the firm adjusts quality to compensate for the negative demand response, re-optimizing its choice of output. The regulator wants to use the payment policy to discipline the firm to truthfully reveal its private information, and not extract any more of an information rent than necessary for incentive compatibility. To this end, the regulator may need to either discipline the firm by setting the price above the first-best level to make it more expensive to produce output, or by setting the price below first-best to decrease the firm's unit revenue. In contrast, when the firm cannot manipulate demand by adjusting quality (because consumer demand is inelastic to quality), then the unit price uniquely determines the quantity demanded. The distortion from the first-best is always downward in this case as the regulator must shade the unit payment in order to extract some of the firm's information rent. This result is also analogous to that found in single unit procurement models because demand is price-inelastic and higher payments always increase the firm's profit.[42]

Before concluding, we highlight two important directions for future research. First, in analyzing the effect of the consumers' incentive response to the contracted unit price we took the extreme position that either the regulator or the consumers are responsible for the entire payment; however, in many regulated markets the government and consumers share responsibility. For example, in voucher programs consumers are provided a voucher for tuition at the school of their choice but schools are not limited to charging the voucher amount and consumers may have to kick in a payment above the voucher. In this way the voucher softens the consumers' price elasticity of demand, but does not make it completely inelastic. Similarly, as a part of the PPACA the government has mandated insurance coverage. To help those for whom premiums would exceed a certain percentage of income, the government provides subsidies, softening the price elasticity of demand for those eligible. Given the prevalence of such mixed payment systems, studying the optimal tiered payment policy is an interesting and important avenue of future research.

[41] Recall that the first-best outcome is generally distorted away from the social optimum as Spence (1975) first showed. This occurs because of the difference in the average consumer's valuation for quality compared to the marginal consumer's valuation. In contrast, the distortions found in this paper are away from the first-best and are a result of the firm's information advantage and ability to alter demand.

[42] For example Baron and Besanko (1984) and Laffont and Tirole (1986).

Secondly, despite assuming a market environment in which the regulator cannot observe the firm's costs, the output, or quality level, the information burden on the regulator is still exceedingly high. The regulator is assumed to know or have a strong prior for the firm's cost of production and the characteristics of consumer demand. Since regulators are generally much less informed about the details of the firm's technology, especially in an environment such as health care where those technologies are evolving rapidly, it will be beneficial for future research to consider the nature of regulatory policies under even more restricted information regimes.

REFERENCES

Aguirre, I. and A. Beitia (2004). Regulating a monopolist with unknown demand: Costly public funds and the value of private information. *Journal of Public Economic Theory 6*(5), 693–706.

Armstrong, M. and D. E. Sappington (2004). Toward a synthesis of models of regulatory policy design with limited information. Working paper.

Auriol, E. (1998). Deregulation and quality. *International Journal of Industrial Organization 16*(2), 169–194.

Baron, D. and D. Besanko (1984). Regulation, asymmetric information, and auditing. *Rand Journal of Economics 15*, 447–470.

Baron, D. and R. Myerson (1982). Regulating a monopolist with unknown costs. *Econometrica 50*, 911–930.

Baron, D. P. (1981). Price regulation, product quality, and asymmetric information. *American Economic Review 71*(1), 212–220.

Beitia, A. (2003). Hospital quality choice and market structure in a regulated duopoly. *Journal of Health Economics 22*, 1011–1036.

Caillaud, B., R. Guesnerie, and J.-J. Tirole (1988). Government intervention in production incentives theory: A review of recent contributions. *RAND Journal of Economics Journal of Economics 19*(1), 1–26.

Chalkley, M. and J. M. Malcomson (1998a). Contracting for health services when patient demand does not reflect quality. *Journal of Health Economics 17*, 1–19.

Chalkley, M. and J. M. Malcomson (1998b). Contracting for health services with unmonitored quality. *Economic Journal 108*, 1093–1110.

Dasgupta, P., P. Hammond, and E. Maskin (1979). The implementation of social choice rules: Some general results on incentive compatibility. *Review of Economic Studies 46*, 185–216.

Ellis, R. P. and T. G. McGuire (1986). Provider behavior under prospective reimbursement. *Journal of Health Economics 5*, 129–151.

Gravelle, H. (1999). Capitation contracts: access and quality. *Journal of Health Economics 18*(3), 315–340.

Guesnerie, R. and J.-J. Laffont (1984). A complete solution to a class of principal-agent problems with an application to the control of a self-managed firm. *Journal of Public Economics 25*, 329–369.

Jack, W. (2005). Purchasing health care services from providers with unknown altruism. *Journal of Health Economics 24*, 73–93.

Laffont, J. and D. Martimort (2002). *The Theory of Incentives: THe Principal-Agent Model.* Princeton University press.

Laffont, J.-J. and J. Tirole (1986). Using cost observation to regulate firms. *Journal of Political Economy 94*, 614–641.

Laffont, J.-J. and J. Tirole (1993). *A Theory of Incentives in Procurement and Regulation.* MIT Press.

Landon, B. E., C.-L. T. Normand, A. Lessler, J. O'Malley, S. Schmaltz, J. L. Loeb, and B. J. McNeil (2006). Quality of care for the treatment of acute medical conditions in u.s. hospitals. *Archives of Internal Medicine 166*, 2511–2517.

Lewis, T. R. and D. E. Sappington (1988). Regulating a monopolist with unknown demand. *American Economic Review 78*(5), 986–998.

Lewis, T. R. and D. E. Sappington (1992). Incentives for conservation and quality-improvement by public utilities. *American Economic Review 82*(5), 1321–1340.

Ma, C.-T. A. (1994). Health care payment systems: Cost and quality incentives. *Journal of Economics and Management Strategy 3*(1), 93–112.

Milgrom, P. (2001). *Putting Auction Theory to Work.* Cambridge University Press.

Myerson, R. (1979). Incentive compatibility and the bargaining problem. *Econometrica 47*, 61–73.

Riordan, M. (1984). On delegating price authority to a regulated firm. *Rand Journal of Economics 15*, 108–115.

Rogerson, W. P. (1987). On the optimality of menus of linear contracts. Northwestern Discussion Paper #714R.

Rogerson, W. P. (1994). Choice of treatment intensities by a nonprofit hospital under prospective pricing. *Journal of Economics and Management Strategy 3*, 7–51.

Sappington, D. E. (1982). Optimal regulation of of research and development under imperfect information. *Bell Journal of Economics 13*, 354–368.

Spence, A. M. (1975). Monopoly, quality, and regulation. *Bell Journal of Economics 6*(2), 417–429.

Wolinsky, A. (1997). Regulation of duopoly: Managed competition vs regulated monopolies. *Journal of Economics and Management Strategy 6*(4), 821–847.

Appendix A: Mathematical Proofs

The following 5 Lemmas characterize further the regulator's problem for a pure output maximizing firm. Most of these lemmas follow Rogerson (1994). First, recall that the firm's optimization problem is defined as

(A-1) $$\max_x \beta\Pi(x; p, T, \theta) + (1-\beta)\phi(x) \text{ subject to } \Pi \geq 0.$$

Let $l(\theta)$ denote the Lagrange multiplier then the firm's FOCs yield:

$$l = \frac{\big(\beta(p-g_x)+(1-\beta)\phi'\big)}{(g_x-p)}, \tag{A-2}$$

$$px - g + T = 0, \tag{A-3}$$

$$0 \leq l \leq \infty. \tag{A-4}$$

The multiplier, l, identifies the shadow price of increasing firm profit in terms of lost community benefit. It is straightforward to show that the shadow price is decreasing in the cost-state ($l_\theta < 0$), which follows because the firm's marginal cost of production is increasing with the cost state, making it more expensive to produce the same quantity. Lastly, when, at the limit, the unit price equals the quality-adjusted marginal cost ($p = g_x$), the Lagrange multiplier will assume the value ∞. It is clear that when the firm's optimal choice of output leaves it with some positive profit, $\beta(p-g_x)+(1-\beta)\phi'\big) = 0$ and $l(\theta) = 0$.

Throughout the analysis we have used the quality-adjusted cost function $g(x;p,\theta) = c(x,q)$. Continuing to use $g(\cdot)$, it is useful to denote AC_{qa} as the quality-adjusted average cost and MC_{qa} as the quality-adjusted marginal cost, formally

$$AC_{qa} = g(x,\theta)/x,$$
$$MC_{qa} = g_x(x,\theta).$$

Because the cost of production is monotonically increasing and convex in x it is easy to show the following lemma.

LEMMA 4: *There exists a unique $x \geq 0$ such that $AC_{qa}(x) = MC_{qa}(x)$.*

PROOF:

This follows from the fact that there exists an x such that $AC(x) = MC(x)$ for all $q \geq 0$ and the MC_{qa} (g_x) is increasing convex in x.
Let x^E denote the unique x satisfying $AC_{qa}(x) = MC_{qa}(x)$ and let x^{fb} be the first-best quantity, then inducing the first-best outcome can be characterized by the relative value of x^{fb} to x^E.

The following lemmas characterize the first-best policy.

LEMMA 5: *When the firm is a pure output-maximizer, a policy inducing the first-best must include a positive lump-sum transfer for all quantities $0 < x^{fb} < x^E$ and $p(\theta) \leq MC_{qa}\big(x^{fb}(\theta); p(\theta), \theta\big)$.*

PROOF:

When $x^{fb} < x^E$ then $MC_{qa}(x^{fb}) > AC_{qa}(x^{fb})$. If $T(\theta) = 0\ \forall\theta$ then $p(\theta) \leq AC_{qa}(x^{fb}; p(\theta), \theta)$ if the firm is to produce x^{fb}. However, $p(\theta) \geq AC_{qa}(x^{fb}; p(\theta), \theta) \Rightarrow p(\theta) \geq MC_{qa}(x^{fb}; p(\theta), \theta)$ and the firm can continue to produce beyond x^{fb}. A positive transfer ($T(\theta) = 0$) functions as a subsidy, lowering the firm's average cost curve and sliding the efficient scale down the marginal cost curve. The transfer must be sufficiently large to insure $p(\theta) \leq MC_{qa}\big(x^{fb}(\theta); p(\theta), \theta\big)$.

LEMMA 6: *When the firm is a pure output-maximizer, the first-best can be induced with only a unit payment, p, when $x^E \leq x^{fb}$.*

PROOF:

When $x^{fb} > x^E$, it must be the case that $MC_{qa}(x^{fb}) < AC_{qa}(x^{fb})$, by definition of x^E. Therefore at $p(\theta) = AC_{qa}\big(x^{fb}; p(\theta), \theta\big)$ the firm will increase quality until output equals x^*.

Lemma 6 corresponds with lemma 3.6 in Rogerson (1994).

LEMMA 7: *First-best can be induced with only a lump-sum transfer, T, for any $x^{fb} > 0$.*

PROOF:

This follows immediately from (A-3). Because the firm's cost is increasing in output, the regulator can induce the firm to output the first-best quantity simply by giving it a lump-sum payment equivalent to the unique cost of producing that output.

To provide insight into the characteristics of the problem permitting the payment rule determined by Eq. (12) to also satisfy incentive compatibility, we decompose the regulator's objective function into two parts. The term $V - g$ represents the social surplus and the term $\lambda H g_\theta$ represents the social loss. The second-best optimal price is the price equating the change in social surplus from raising the price to the change in social loss; i.e., p^{sb} solves $d\{V - g\}/dp = \lambda H d\{g_\theta\}/dp$. Let $MSB \equiv d\{V - g\}/dp$ denote the marginal increase to the social surplus from raising the price and let $MSC \equiv \lambda H d\{g_\theta\}/dp$ denote the marginal increase to the social cost. If $dMSB/d\theta > dMSC/d\theta$, then the increase in the benefit exceeds the loss and it follows that the regulator will prefer to increase the price with an increase in the cost state; i.e., $dp^{sb}/d\theta > 0$. Otherwise the increase in social loss exceeds the increase in social benefit and the regulator prefers to decrease the price with an increase in the cost state.

Similarly, the change in the firm's value due to increasing the price by dp can be decomposed into the marginal benefit $x + (1-\beta)/\beta\phi'$ and the marginal cost g_p. The firm's profit is strictly concave in p therefore, if $x^* > g_p$, then the firm's profit is increasing in price, otherwise, it is decreasing. More importantly, however, if $dx^*/d\theta > \partial^2 g/\partial\theta\partial p$, then the firm prefers a higher price for higher values of the state parameter. The following lemma follows from the preceding discussion and is essentially a restatement of property *(ii)* of lemma 1.

LEMMA 8: *The p^{sb} solving (12) is incentive compatible only if*

$$\frac{dMSB}{d\theta} > \frac{dMSC}{d\theta} \quad \textit{whenever} \quad \frac{dx^*}{d\theta} > \frac{\partial^2 g}{\partial\theta\partial p}.$$

PROOF OF PROPOSITION 2

For any given p, the firm chooses the x^* that sets $p = g_x - \big((1-\beta)/\beta\big)\phi'$. Thus, the socially preferred output is the output setting $p = V_x - \big((1-\beta)/\beta\big)\phi'$. Because of the concavity of the regulator's problem, if $p < V_x - \big((1-\beta)/\beta\big)\phi'$ for any p, then $x^* < \arg\max_x(V - g)$, and if $p > V_x - \big((1-\beta)/\beta\big)\phi'$ for any p, then $x^* > \arg\max_x(V - g)$.

PROOF OF LEMMA 1

Condition *(i)* is a necessary condition for an optimum and is derived as follows. First we derive the condition when $\Pi(x^*) > 0$ and the firm has a mixed objective.

Recall the firm's value function is defined as

$$\text{(A-5)} \qquad \mathcal{U}(\hat\theta,\theta) = \beta\big[p(\hat\theta)x^*\big(p(\hat\theta),\theta\big) - g\big(x^*\big(p(\hat\theta),\theta\big),\theta\big) + T(\hat\theta)\big] + (1-\beta)\phi\big(x^*\big(p(\hat\theta),\theta\big)\big).$$

A necessary condition for truth-telling is that the announcement of θ results in maximal profit. The first-order condition for truth-telling is thus

$$\frac{\partial\mathcal{U}}{\partial\hat\theta}(\hat\theta,\theta) = \beta\left[\frac{dp}{d\hat\theta}x^*\big(p(\hat\theta),\theta\big) + p(\hat\theta)\frac{\partial x^*}{\partial p}\frac{dp}{d\hat\theta} - \frac{\partial g}{\partial x^*}\frac{\partial x^*}{\partial p}\frac{dp}{d\hat\theta} - \frac{\partial g}{\partial p}\frac{\partial p}{\partial\hat\theta} + \frac{dT}{d\hat\theta}\right] + (1-\beta)\frac{d\phi}{dx}\frac{\partial x^*}{\partial p}\frac{dp}{d\hat\theta} = 0.$$

Applying the envelope theorem to the first-order condition of $\mathcal{U}(\theta) = \mathcal{U}(\theta, \theta)$ implies

$$\frac{d\mathcal{U}}{d\theta}\big(\hat{\theta}(\theta), \theta\big) = \frac{\partial U(\theta)}{\partial \theta} = \frac{\partial U}{\partial \hat{\theta}} \frac{\partial \hat{\theta}}{\partial \theta} + \frac{\partial U}{\partial x^*} \frac{\partial x^*}{\partial \theta} - \beta \frac{\partial g}{\partial \theta},$$

where $\hat{\theta}(\theta)$ is the firm's announcement strategy given the true demand state is θ; i.e. $\hat{\theta} : \Theta \to \Theta$. Thus, by applying the envelope theorem, a necessary condition for the optimal payment policy is that

$$\frac{d\mathcal{U}}{d\theta} = -\beta \frac{\partial g}{\partial \theta}. \tag{A-6}$$

When $\Pi(x^*) = 0$ and the firm is a pure output-maximizer then x^* solves $px^* - g(x^*; p, \theta) + T = 0$. Because the firm leaves itself with 0 profits for any state θ, we have

$$\frac{d\mathcal{U}}{d\theta} = \beta \frac{d\Pi}{d\theta} + (1-\beta)\frac{d\phi}{d\theta} = (1-\beta)\frac{d\phi}{dx}\frac{dx^*}{d\theta}.$$

Using the implicit function theorem, $\frac{dx^*}{d\theta}$ can be expressed as

$$\frac{dx^*(\hat{\theta}(\theta), \theta)}{d\theta} = -\frac{p_\theta \frac{d\hat{\theta}}{d\theta} x - g_p p_\theta \frac{d\hat{\theta}}{d\theta} - g_\theta + T_\theta \frac{d\hat{\theta}}{d\theta}}{p - g_x} = \frac{dx^*}{d\hat{\theta}}\frac{d\hat{\theta}}{d\theta} - \frac{-g_\theta}{p - g_x}. \tag{A-7}$$

The firm announces the $\hat{\theta}$ which maximizes output, therefore by the envelope theorem the first term on the RHS of (A-7) is zero, yielding:

$$\frac{d\mathcal{U}}{d\theta} = \frac{\partial \mathcal{U}}{\partial \theta} = -(1-\beta)\phi'\big(x^*(p(\theta), \theta)\big)\left[\frac{g_\theta\big(x^*(p(\theta), \theta); p(\theta), \theta\big)}{g_x\big(x^*(p(\theta), \theta); p(\theta), \theta\big) - p(\theta)}\right].$$

Next, condition *(ii)* of the lemma represents a sufficient condition. To show sufficiency when the firm's choice of output leaves it with positive profit, we apply the envelope theorem to (A-5) yielding:

$$\frac{\partial \mathcal{U}(\hat{\theta}, \theta)}{\partial \hat{\theta}} = \beta\left[\frac{dp}{d\hat{\theta}}x - \frac{\partial g}{\partial p}\frac{dp}{d\hat{\theta}} + \frac{dT}{d\hat{\theta}}\right]. \tag{A-8}$$

From the fact that $\left.\frac{\partial \mathcal{U}(\hat{\theta}, \theta)}{\partial \hat{\theta}}\right|_{\hat{\theta}=\theta} = 0$ we have

$$\frac{dT}{d\hat{\theta}} = \frac{\partial g}{\partial p}\frac{\partial p}{\partial \hat{\theta}} - \frac{dp}{d\hat{\theta}}x\big(p(\hat{\theta}), \hat{\theta}\big). \tag{A-9}$$

Plugging (A-9) into (A-8) yields

$$\begin{aligned}\frac{\partial \mathcal{U}(\hat{\theta}, \theta)}{\partial \hat{\theta}} &= \beta\frac{dp}{d\hat{\theta}}\Big[\Big(x\big(p(\hat{\theta}), \theta\big) - g_p\big(x(p(\hat{\theta}), \theta); p(\hat{\theta}), \theta\big)\Big) - \Big(x\big(p(\hat{\theta}), \hat{\theta}\big) - g_p\big(x(p(\hat{\theta}), \hat{\theta}); p(\hat{\theta}), \hat{\theta}\big)\Big)\Big] \\ &= \beta\frac{dp}{d\hat{\theta}}\big[\mathcal{U}_p(\hat{\theta}, \theta) - \mathcal{U}_p(\hat{\theta}, \hat{\theta})\big].\end{aligned}$$

By the intermediate value theorem there exists a $\tilde{\theta} \in [\theta, \hat{\theta}]$ if $\theta < \hat{\theta}$ or $\tilde{\theta} \in [\hat{\theta}, \theta]$ if $\theta > \hat{\theta}$ such that

$$\text{(A-10)} \qquad \frac{\partial U(\hat{\theta}, \theta)}{\partial \hat{\theta}} = \beta \left[\frac{dp}{d\hat{\theta}} \frac{\partial^2 \mathcal{U}(\hat{\theta}, \tilde{\theta})}{\partial p \partial \tilde{\theta}} (\theta - \hat{\theta}) \right].$$

Because $U_T = \beta$, the second-order cross partial derivative of Eq. (A-10), $\mathcal{U}_{\theta p}$, is equal to $\frac{\partial}{\partial \theta}(U_p/U_T)$. The condition $sign[dp/d\theta] = sign\left[\frac{\partial}{\partial \theta}(U_p/U_T)\right]$ implies

$$\frac{\partial \mathcal{U}(\hat{\theta}, \theta)}{\partial \hat{\theta}} \geq 0 \text{ when } \hat{\theta} < \theta$$
$$\frac{\partial \mathcal{U}(\hat{\theta}, \theta)}{\partial \hat{\theta}} \leq 0 \text{ when } \hat{\theta} > \theta$$

Thus, $\hat{\theta} = \theta$ is a global maximizer and the payment policy induces truthful revelation when $sign[dp/d\theta] = sign\left[\frac{\partial}{\partial \theta}(U_p/U_T)\right]$.

To prove condition *(ii)* for a pure output-maximizing firm we employ a slightly different approach. Incentive compatibility is satisfied if and only if the firm maximizes its output with a truthful announcement of the state. Therefore incentive compatibility is satisfied for an output-maximizing firm if and only if for any θ_1 and θ_2 in Θ where $\theta_1 < \theta_2$, the following hold

$$\text{(A-11)} \qquad x^*\big(p(\theta_2), T(\theta_2), \theta_1\big) \leq x^*\big(p(\theta_1), T(\theta_1), \theta_1\big),$$
$$\text{(A-12)} \qquad x^*\big(p(\theta_1), T(\theta_1), \theta_2\big) \leq x^*\big(p(\theta_2), T(\theta_2), \theta_2\big).$$

Adding (A-11) and (A-12) gives

$$x^*\big(p(\theta_2), T(\theta_2), \theta_2\big) - x^*\big(p(\theta_1), T(\theta_1), \theta_2\big) \geq x^*\big(p(\theta_2), T(\theta_2), \theta_1\big) - x^*\big(p(\theta_1), T(\theta_1), \theta_1\big),$$

implying

$$\text{(A-13)} \qquad \int_{\theta_1}^{\theta_2} \int_{\theta_1}^{\theta_2} \frac{d^2 x^*}{d\hat{\theta} d\theta} d\hat{\theta} d\theta \geq 0.$$

Because (A-13) is true for all $\theta_1, \theta_2 \in \Theta$ it implies $\frac{d^2 x}{d\hat{\theta} d\theta} \geq 0$, which is equivalent to

$$\text{(A-14)} \qquad \frac{d^2 x^*}{dT d\theta} \frac{dT}{d\hat{\theta}} + \frac{d^2 x^*}{dp d\theta} \frac{dp}{d\hat{\theta}} \geq 0.$$

We can simplify (A-14) by observing that a truthful announcement of the state parameter is optimal if

$$\text{(A-15)} \qquad \left.\frac{dx^*}{d\hat{\theta}}\right|_{\hat{\theta}=\theta} = \left.\frac{dx^*}{dp} \frac{dp}{d\hat{\theta}}\right|_{\hat{\theta}=\theta} + \left.\frac{dx^*}{dT} \frac{dT}{d\hat{\theta}}\right|_{\hat{\theta}=\theta} = 0.$$

Using (A-15) we can rewrite (A-14) as

$$\frac{\partial}{\partial\theta}\left(\frac{dx^{*}/dp}{dx^{*}/dT}\right)\frac{dp}{d\hat{\theta}}\bigg|_{\hat{\theta}=\theta}\geq 0.^{43} \tag{A-16}$$

Eq. (A-16) is a special case of the condition derived in Theorem 1 of Guesnerie and Laffont (1984). The term $(dx^{*}/dp)/(dx^{*}/dT) = (dU/dp)/(dU/dT)$ is the output-maximizing firm's MRS of unit payment for fixed transfer so the firm's objective function satisfies the SCP when $d\{(dU/dp)/(dU/dT)\}/d\theta$ is monotonic for all $\theta \in \Theta$.

PROOF OF COROLLARY 1

Starting with the definition of $\mathcal{U}(\theta)$ and property *(i)* of Lemma 1 we have:

$$\begin{aligned}\mathcal{U}(\theta) &= \mathcal{U}(\overline{\theta}) + \beta\int_{\theta}^{\overline{\theta}} g_{\theta}(x^{*};p,\theta)d\theta \\ &= (1-\beta)\varphi(\overline{\theta}) + \beta\int_{\theta}^{\overline{\theta}} g_{\theta}(x^{*};p,\theta)d\theta. \end{aligned} \tag{A-17}$$

Substituting $\mathcal{U}(\theta) = \beta\Pi(\theta) + (1-\beta)\varphi(\theta)$ in (A-17), rearranging, and taking expectations gives:

$$E\Pi = \int_{\underline{\theta}}^{\overline{\theta}}\left\{\tfrac{1-\beta}{\beta}\big(\varphi(\overline{\theta})-\varphi(\theta)\big) + \int_{\theta}^{\overline{\theta}} g_{\theta}(x^{*};p,\theta)d\theta\right\}dF(\theta).$$

Integrating by parts yields the expression:

$$E\Pi = \int_{\underline{\theta}}^{\overline{\theta}}\left\{\tfrac{1-\beta}{\beta}\big(\varphi(\overline{\theta})-\varphi(\theta)\big) - \tfrac{F(\theta)}{f(\theta)}g_{\theta}(x^{*};p,\theta)\right\}dF(\theta).$$

PROOF OF PROPOSITION 3

Because the regulator's problem is quasiconcave in x and p (see footnote 28) we have $d^{2}\{V-g\}(x;p) = d^{2}\{V-g\}(p)/dp^{2} < 0$. The first-best price, p^{fb} is the price solving $d\{V-g\}(p^{fb})/dp = 0$. When $d\{g_{\theta}(x^{*};p,\theta)\}/dp = 0$ for all $\theta \in \Theta$ the firm extracts no rents and from Eq. (12) it is clear that p^{sb} solves $d\{V-g\}(p)/dp = 0$. Therefore $p^{sb} = p^{fb}$. When $d\{V-g\}(p^{sb})/dp > 0$ concavity in the regulator's problem implies $p^{sb} < p^{fb}$ and when $d\{V-g\}(p^{sb})/dp < 0$ concavity implies $p^{sb} > p^{fb}$. From Eq. (12) , $sign[d\{V-g\}(p^{sb})/dp]$ depends on whether the firm's information rents are increasing or decreasing with the unit payment. Therefore, when the firm's information rents are decreasing with the unit payment then $d\{g_{\theta}(x^{*};p,\theta)\}/dp < 0$ and we have $p^{sb} > p^{fb}$ and when the firm's information rents are increasing then $d\{g_{\theta}(x^{*};p,\theta)\}/dp > 0$ and we have $p^{sb} < p^{fb}$.

PROOF OF PROPOSITION 4 The proof follows immediately from Proposition 3 and from identifying $sign[d\{g_{\theta}\}/dp]$. There is no distortion in the price at $\underline{\theta}$ since $F(\underline{\theta}) = 0$ and output is not distorted since it is uniquely determined by the unit price.

[43]This condition can equivalently be written as

$$\frac{\partial}{\partial\theta}\left(\frac{\partial x^{*}/\partial T}{\partial x^{*}/\partial p}\right)\frac{dT}{d\hat{\theta}}\bigg|_{\hat{\theta}=\theta}\geq 0.$$

PROOF OF PROPOSITION 5

The conditions of Lemma 1 identify the necessary and sufficient conditions for an incentive compatible payment policy that must be satisfied by any set of payment rules: $\{p^{np}(\theta), T^{np}(\theta)\}_{\theta\in\Theta}$. The unit payment that induces the first-best level of output is derived by setting $dx^* /d\theta = dx^{fb} /d\theta$, where $dx^* /d\theta$ is defined in Lemma 1 and solving for p. By the integral form of the envelope theorem, we have $x^*(p(\hat{\theta}), \theta) = x^{fb}(\bar{\theta}) - \int_{\theta}^{\bar{\theta}} (\partial x^{fb}/\partial\tilde{\theta})d\tilde{\theta}$, if and only if $x^* = x^{fb}$ at every $\theta \in \Theta$, thus a payment policy that induces $dx^* /d\theta = dx^{fb} /d\theta$ and sets $x^*(\bar{\theta}) = x^{fb}(\bar{\theta})$ induces the first-best output in every state.

From Lemma 5, the unit payment must always be less than or equal to the marginal cost at the induced quantity. Because $g_\theta > 0$, this condition requires that the first-best output be weakly decreasing with the cost-state, $dx^{fb} /d\theta \leq 0$, otherwise, $p^{np}(\theta) > g_x\big(x^*(\theta); p(^{np}(\theta), \theta\big)$ and the firm is not left with zero profit. Consequently it will choose a higher output and the payment rule does not induce the first-best output. It must also be the case that $p(\theta) \neq g_x\big(x^*(\theta); p(\theta), \theta\big)$ for all $\theta \in \Theta$, otherwise $dx^* /d\theta$ is not bounded at some $\theta \in \Theta$, hence not absolutely continuous, and the envelope theorem cannot apply. The assumptions on the value and cost functions insure this cannot happen. First, the regulator's problem insures that $p^{np}(\theta)$ is unique for every state θ; and second, the firm's problem insures that x^* is unique for every p. Therefore, for a given state the first-best quantity is unique and there does not exist any such $\theta' \in \Theta$ such that $\lim_{\theta\to\theta'} dx^{fb}(\theta) /d\theta = \infty$.

Finally, the unit payment must be restricted to a nonnegative value because demand is not defined outside the domain $p \geq 0$ and the pricing rule does not guarantee that the price will be nonnegative. Putting the three conditions of the proposition together ($dx^{fb} /d\theta < 0$, $dp^{np} /d\theta < 0$, and $p^{np}(\theta) \geq 0$) imply that the payment rule reported by the proposition is a proper, incentive compatible payment rule that will induce the first-best output, and any payment rule that induces the first-best output must satisfy the three conditions of the proposition.

PROOF OF PROPOSITION 6

The regulator's problem is complicated by the fact that, in addition to the firm's choice of output, the consumers' demand is also a function of the unit price, p. Inducing the consumers to demand the first-best quantity requires that p^{np} solve

$$d\{V - p\} /dp = 0. \tag{A-18}$$

Inducing the firm to supply the appropriate level of quality while maintaining incentive compatibility requires that p^{np} also satisfy

$$p^{np}(\theta) = g_x\big(x^{fb}; p^{np}(\theta), \theta\big) + \frac{g_\theta\big(x^{fb}; p^{np}(\theta), \theta\big)}{dx^{fb} /d\theta} \text{ for all } \theta \in \Theta. \tag{A-19}$$

Eq. (A-18) and (A-19) are independent, thus $p^{np}(\theta) = p^{fb}(\theta)$ for all $\theta \in \Theta$ only if $dp^{np} /d\theta = dp^{fb} /d\theta$ at all $\theta \in \Theta$. It is clear that, generically,

$$\frac{dp^{fb}}{d\theta} = -\left(\frac{d^2\{V-g\}}{dpd\theta} \bigg/ \frac{d^2\{V-g\}}{dp^2}\right) \neq \frac{dp^{np}}{d\theta}.$$

Hence, $p^{np} \neq p^{fb}$ and Eq. (A-18) is not equal to zero at all $\{x^{fb}(\theta), p^{np}(\theta), T^{np}(\theta)\}_{\theta\in\Theta}$ given the strict concavity of the regulator's problem. Therefore $\{x^{fb}(\theta), p^{np}(\theta), T^{np}(\theta)\}_{\theta\in\Theta}$ is not first-best optimal for a marketed good.

PROOF OF LEMMA 2 Taking derivatives of the firm's profit with respect to price and transfer payment gives $\mathcal{U}_p/\mathcal{U}_T = x^*$. From the conjugate pairs theorem $sign[dx^*/d\theta] = sign[\mathcal{U}_{x\theta}]$ and $\beta\Pi_{x\theta} = -\beta g_{x\theta} < 0$.

PROOF OF LEMMA 3

The proof for Lemma 1 continues to hold if $g_p = 0$ so applies to a nonmarketed good as well. For a nonmarketed good the SCP holds and is negative for all class of functions satisfying the model's properties. Therefore the condition $sign[dp/d\theta] = sign\left[\frac{\partial}{\partial\theta}(U_p/U_T)\right]$ requires $sign[dp/d\theta] < 0$.

PROOF OF PROPOSITION 7

The payment policy for a mixed-objectives firm is derived by taking the FOC of the regulator's problem and substituting in the firm's first-order condition. Combining Lemmas 5 - 7 yields the pricing rule for the out-put maximizing firm.

PROOF OF PROPOSITION 9

The FOC of the Hamiltonian must satisfy:

$$\begin{aligned}\tfrac{d}{dp}\left\{\tfrac{d}{d\theta}\{V-g\}\right\} &= \tfrac{d}{dp}\left\{(V_x - g_x)\tfrac{dx^*}{d\theta}\right\} = 0,\\ &= \tfrac{d}{d\theta}\left\{(V_x - g_x)\tfrac{dx^*}{dp}\right\} = 0.\end{aligned}$$

Where $V_p = g_p = 0$ since there is no demand response to price. The FOC can be manipulated to isolate $dx^*/d\theta$; that is, rearranging the FOC yields:

$$\text{(A-20)} \qquad \frac{dx^*}{d\theta} = -\frac{\frac{d}{d\theta}\left\{(V_x - g_x)\frac{dx^*}{dp}\right\}}{\frac{d}{dx}\left\{(V_x - g_x)\frac{dx^*}{dp}\right\}} = -\frac{\frac{d}{d\theta}\left\{\frac{d}{dp}(V-g)\right\}}{\frac{d}{dx}\left\{\frac{d}{dp}(V-g)\right\}}.$$

The RHS of (A-20) is the definition of $dx^{fb}/d\theta$, therefore if the payment policy induces $dx^*/d\theta = dx^{fb}/d\theta$ then the first-order necessary condition for maximization is satisfied. Because the regulator's problem is quasi-concave the condition is also sufficient and the regulator achieves the first-best outcome.

Appendix B: Asymmetric Information about Demand (Not For Publication)

Lewis and Sappington (1988) find in a model without quality that asymmetric demand information does not result in an output distortion away from the first-best levels, in stark contrast to when asymmetric information is with respect to the firm's cost. We investigate whether the present results are robust to the source of asymmetric information.

To see how asymmetric demand information alters the results we make the following changes to the model. First, the state parameter now represents a shock to demand instead of cost: $x(q, p; \theta)$. To maintain consistency with the model with asymmetric knowledge of cost, we assume that higher states result in *less* demand at the same quality level and unit price: $x_\theta < 0$. As before we will work with the quality demand function $q(x, p; \theta)$; therefore, because $x_\theta < 0$, it must be the case that $q_\theta > 0$ and in higher demand states the firm's cost of production is higher for the same equilibrium quantity.

It will continue to be more convenient to reduce the problem by one dimension and work with quality-adjusted cost and social value functions. The quality-adjusted cost function is now defined as:

$$g(x; p, \theta) = c\big(x, q\big(x, p; \theta\big)\big).$$

Because higher demand-states *soften* demand, the partial derivatives and cross-partials are consistent across models: $g_x > 0$, $g_p > 0$, $g_\theta > 0$, and $g_{\theta x} \geq 0$.

The quality-adjusted social value function now takes the demand state as an argument,

$$V(x; p, \theta) = B\big(x, q(x, p; \theta)\big),$$

as it is now dependent upon the demand state. A higher state results in a *left-ward* shift in the demand curve so the social value to consuming x units must be lower in higher demand-states (i.e., $V_\theta < 0$), and higher in low demand-states.

Condition *(i)* of Lemma 1 identifies a necessary condition for an incentive compatible payment rule with asymmetric knowledge of cost. When the firm earns some information rent at an optimum ($\Pi > 0$), the condition, $dU/d\theta = -g_\theta(x; p, \theta) < 0$, applies regardless of the source of asymmetric information. Similarly, condition *(ii)* identifies a sufficient condition, which is dependent on the properties of the firm's objective function. The properties of g do not change based on the source of information asymmetry, so this condition applies to both cases of asymmetric information. In consequence, the firm earns no rents in the highest state, $\overline{\theta}$, and there is no output distortion in the lowest state, $\underline{\theta}$, regardless of the source of information asymmetry.

For a marketed good, the conditions that determine the relative size of the second- to first-best prices as enumerated by Proposition 4, are independent of the source of asymmetric information. Because the relationship between outputs is depends on how the unit payment affects the firm's information rents, the source of asymmetric information can affect the direction of an output distortion. That is, the relationship between the second- and first-best outputs is dependent on the sign for $d\{g_\theta\}/dp$, which varies with the source of asymmetric information. This can be seen by decomposing $d\{g_\theta\}/dp$ into its component parts:

$$d\{g_\theta(x^*; p, \theta)\}/dp = g_{\theta p} + g_{\theta x}(dx^*/dp).$$

Regardless of the source of information asymmetry, $g_{\theta x} > 0$, and dx^*/dp is unrestricted; however, $sign[g_{\theta p}]$ is somewhat dependent on the source of asymmetric information. When the asymmetry is in cost, $g_{\theta p} > 0$, but when it is in demand, the sign is ambiguous. This follows because with asymmetric knowledge of demand $g_{\theta p} = c_q q_{\theta p}$ and $sign[q_{\theta p}]$ is unrestricted. When $q_{\theta p} > 0$, demand is more sensitive to the price in softer demand states, thus requiring ever increasing levels of quality to compensate for the demand response to an increase in price; and, when $q_{\theta p} < 0$ demand is less sensitive to price in softer demand states, reducing the rents the firm can extract, $g_{\theta p} < 0$.

When the asymmetry is with cost, then $g_{\theta p} > 0$ and the partial change in the firm's rent with price and the partial change in the firm's rent with output move in the same direction, i.e., $g_{\theta p}/g_{\theta x} > 0$. With asymmetric demand information the two may move counter to one another. When $g_{\theta p}/g_{\theta x} > 0$ then the relationship between first- and second-best prices and output are determined by Proposition 4. Proposition 10 identifies the relationship between first- and second-best prices when $g_{\theta p}/g_{\theta x} < 0$.

PROPOSITION 10: *With asymmetric knowledge of demand, if $g_{\theta p}/g_{\theta x} > 0$, then the relative size of the second- to first-best price and output is identical to when the regulator's uncertainty is in cost. Otherwise, the relative size of the second- to first-best price and output is determined by the rules:*

$$\begin{array}{rlll} (i) & 0 < -(g_{\theta p}/g_{\theta x}) < dx^* /dp & \Rightarrow d\{g_\theta\} /dp > 0 \Rightarrow p^{sb}(\theta) < p^{fb}(\theta) & \Rightarrow x^{sb}(\theta) < x^{fb}(\theta), \\ (ii) & 0 < dx^* /dp < -(g_{\theta p}/g_{\theta x}) & \Rightarrow d\{g_\theta\} /dp < 0 \Rightarrow p^{sb}(\theta) > p^{fb}(\theta) & \Rightarrow x^{sb}(\theta) > x^{fb}(\theta), \\ (iii) & dx^* /dp < 0 < -(g_{\theta p}/g_{\theta x}) & \Rightarrow d\{g_\theta\} /dp < 0 \Rightarrow p^{sb}(\theta) > p^{fb}(\theta) & \Rightarrow x^{sb}(\theta) < x^{fb}(\theta), \end{array}$$

for all $\theta \in [\underline{\theta}, \overline{\theta})$ and $p^{sb}(\overline{\theta}) = p^{fb}(\overline{\theta})$.

The intuition behind Proposition 10 follows similarly to Proposition 4. In case *(i)*, $dx^* /dp > -(g_{\theta p}/g_{\theta x})$ and the second-best output must be under-supplied relative to first-best because the firm's rents are increasing with the unit payment and the regulator will set a price below the first-best in order to limit the firm's rents. Furthermore, because price and output are below first-best levels, quality must be unambiguously lower than the first-best level of quality. In case *(ii)*, the firm still optimally increases output with a price increase, but the firm's rents are decreasing faster with the change in output than they are increasing with a change in price resulting in a net decrease in rents with an increase in price. To limit the firm's rents, the regulator will set a price above first-best, resulting in an oversupply. Moreover, because prices and output are above first-best levels, then it is clear that quality is unambiguously above the first-best level as well. Similarly, the firm's information rent is decreasing in the price in case *(iii)*, the difference is the firm optimally chooses a lower output quantity with a higher unit price resulting in an undersupply relative to first-best.

Lastly, because g_θ is independent of the source of information asymmetry, all of the results remain unchanged when the firm's preference for community benefit is so strong that it is a pure output-maximizer. Reflecting this, the proofs include a parameterization that generically represents either asymmetric information for cost or demand. Notably the reason why Lewis and Sappington (1988) find that asymmetric information in demand does not lead to an output distortion is precisely because $g_\theta(\cdot) = c_\theta(\cdot) = 0$ in their model.

Appendix C: An Example (Not For Publication)

To help illustrate the findings of the paper, we present a simple example. What Proposition 4 tells us is that when $sign[dU(\theta)/dp] = sign[dx^*(\theta)/dp]$ then the second-best output will be undersupplied relative to first-best, and when the two derivatives have opposite signs, second-best will be oversupplied. In order to provide an example that illustrates this finding we start with a simple, parameterized demand function

$$x(q,p) = \beta + q - \alpha p^2,$$

where $\alpha > 0$ and $\beta > 0$. The quality-demand function is thus $q(x,p) = x - \beta + \alpha p^2$. For the firm's cost of production consider the function

$$c(x,q;\theta) = \theta x q.$$

By plugging $q(x,p)$ into $c(\cdot)$ we can define the quality-adjusted cost function as

$$g(x;p,\theta) = \theta\big(x^2 + x(\alpha p^2 - \beta)\big).$$

The quality-adjusted cost-function produces the following partial derivatives:

$$\begin{array}{lll} g_\theta = x^2 + x(\alpha p^2 - \beta) & g_x = \theta(2x + \alpha p^2 - \beta) & g_p = 2\theta\alpha x p \\ g_{\theta p} = 2\alpha x p & g_{xx} = 2\theta & g_{px} = 2\theta\alpha p \\ g_{\theta x} = 2x + \alpha p^2 - \beta & & \end{array}$$

Notice that the quality adjusted cost is increasing in θ and strictly convex in x and p. Given a payment policy $\{p(\theta), T(\theta)\}_{\theta\in\Theta}$ and the quality-adjusted cost function g, the firm's optimization program can be expressed as

$$\max_x p(\theta)x - g\big(x;p(\theta),\theta\big) + T(\theta).$$

Taking the first order condition yields the firm's maximizer:

$$x^*(p,\theta) = \tfrac{1}{2\theta}(p(1-\theta\alpha p) + \theta\beta).$$

The reason we chose the particular form of the demand function is because dx^*/dp may either be positive or negative depending on the values of the parameters:

$$\frac{dx^*}{dp} = \begin{cases} > 0 & \text{when} \quad \theta\alpha p < 1, \\ < 0 & \text{when} \quad \theta\alpha p > 1. \end{cases}$$

Let the regulator's value function be $\ln(x+q)$ and θ be a uniformly distributed variable with supports $\Theta = [2,3]$. When $\alpha = 1/2$ then $dx^*/dp < 0$ and the relationship of the second- to first-best price and quantity follows the top panel of table C1. As the table reports, the second-best price is less than the first-best price and this occurs because the firm's information rent is increasing in the price at the first-best. Second, the table reports that the second-best output is higher than the first-best output. This occurs because the firm's best response to a price increase at the second-best price is to decrease output. Because the second-best price is less than the first-best price, this results in a higher output than would be attained at the first-best price level.

Table C1—Price and Quantity Differences

α		2.0	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0
$\frac{1}{2}$	$\%\Delta p$:	0.00	-1.44	-2.69	-3.77	-4.72	-5.57	-6.33	-7.01	-7.64	-8.21	-8.73
	$\%\Delta x$:	0.00	0.17	0.32	0.45	0.57	0.68	0.78	0.86	0.94	1.01	1.07
$\frac{1}{10}$	$\%\Delta p$:	0.00	-1.95	-3.59	-4.97	-6.15	-7.16	-8.04	-8.80	-9.47	-10.06	-10.57
	$\%\Delta x$:	0.00	-0.38	-0.65	-0.84	-0.97	-1.05	-1.09	-1.11	-1.11	-1.09	-1.07

When $\alpha = 1/10$ then $dx^* /dp > 0$. In this case the firm's information rents are still increasing with the price causing the second-best price to fall below the first-best. However, because the firm's best response to a price increase is to increase output, this causes the second-best output level to also fall below the first-best output.

The reason α has such a large effect on the outcome is because it serves to increase the price elasticity of demand as it increases. Thus for sufficiently large α it becomes too costly for the firm to maintain some level of demand when the price is increased, whereas for smaller α the increase is revenue is much stronger than the cost of maintaining a high level of demand and the firm optimally chooses to increase the quantity demanded as prices increase.