Regulating a Monopolist with Unverifiable Quality*

Kevin E. Pflum†

September, 2010

Abstract

I analyze the optimal regulatory policies for a monopolist with superior knowledge of either cost or demand and demand is affected by unverifiable quality. In this setting of unobservable output and costs, the regulator is restricted to using a two-part tariff as policy instrument. The optimal payment policies for a marketed and nonmarketed good and a for-profit and not-for-profit firm are derived and compared. I show how the strategic responses to the regulatory policy of both the consumers and firm determine the degree of distortion from the social optimum. The firm's ability to manipulate demand results in distortions above and below the first-best levels. However, a firm's informational advantage may be completely attenuated with a nonmarketed good and not-for-profit firm but with a loss an efficiency from utilizing public funds.

JEL Codes: H42, H57, L20, L31, L50

Key Words: regulating quality, asymmetric demand information, asymmetric cost information, contracting, adverse selection, moral hazard

1 Introduction

Recent efforts by the U.S. Congress to reform the market for health insurance have approached regulating health insurers as public utilities.[1] Treating health insurers or providers as public utilities is appealing due to the large body of knowledge and experience in their regulation. However, the regulatory approaches taken for public utilities may not be completely transferable to health markets. For example, in telecommunications quality can be identified by a quantitative measure such as the time to connect or 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. Similar limitations to regulating quality 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 quality. However, reflecting the difficulty in quantifying quality levels the handbook reports the following (emphasis added):

*I am grateful to Paul J. Healy, Dan Levin, Lixin Ye, Huanxing Yang, Michael Sinkey, Paula Cordero Salas, and the participants of the Ohio State Micro Theory Seminar for helpful discussions and comments.

†Address: Department of Economics, 1945 N. High St., Columbus, OH, 43210. Email: pflum.4@osu.edu

[1] The recently passed Patient Protection and Affordable Care Act contains strong language regarding the ratemaking and degree of coverage for health insurers participating in health exchanges.

> 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.

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.

Mirroring issues inherent to health and education markets, we analyze the design of regulatory policy in a market in which the contracted firm has superior knowledge of some aspect of the market (cost or demand), and in which demand is a function of the unverifiable quality level[2] supplied by the firm. Reflecting the notion that it is difficult to establish what constitutes a unit of service (e.g., treatment quantity or amount of education), the quantity of output is also not observable by the regulator, and hence not contractable. Furthermore, because many regulated markets contain a mix of for-profit and not-for-profit firms[3] we consider how the conduct of the firm affects the nature of the contract and the outcomes attainable by the regulator within the same information set. Thus, the objectives of this study are to first identify and characterize what outcomes the regulator can achieve with such strong information limitations, and second, to provide deeper insights into how the combination of moral hazard and adverse selection in this market scenario affect the regulator's problem by exploring how altering the consumers' access to the good and the firm's market conduct affects these outcomes.

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 et al. (1988), we refer to the good in the former scenario as a *marketed good*, and in the latter as a *nonmarketed good*. The marketed good represents the classical regulatory environment generally associated with public utilities.[4] The nonmarketed good represents the regulatory environment most often attributed to markets for health care,[5] 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).

This paper's findings generate insights into the regulation of firms in the presence of asymmetric information on two levels. First, constraining ourselves to a regulatory environment containing a marketed good and profit-maximizing firm to facilitate comparison with the earlier literature, the results of the model show that one cannot predict *a priori* what form the output distortion will

[2]The firm's choice of quality is systemic in that it cannot vary the level of quality on a consumer-by-consumer basis, nor can it offer its good or service at multiple levels of quality. This reflects the fact that a school may hire better teachers or put into place more effective curricula, or a hospital may have policies that are more effective at reducing hospital-based infections.

[3]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).

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

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

take when the firm's choice of non-contractable quality influences demand. Depending on the relative price- and quality-elasticities of demand, we find that the second-best contract may result in a distortion of output that may be either above *or* below the socially-optimal levels. This is in marked contrast to similar models which do not include quality (implicitly setting the quality-elasticity of demand to zero) that result in the familiar output distortions below the first-best level. However, the distortion away from the social optimum is similar to earlier studies when the good is nonmarketed because the price-elasticity of demand is forced to zero and output is uniquely determined by the quality level alone, and thus distorted strictly downward.

The second general insight of the model is that the regulator need not find ways to reduce information asymmetries between it and the firm in order to reduce the distortions caused by the firm's informational advantage. Instead, the regulator can analyze how the firm and consumers interact with its available policy instruments and, if possible and beneficial, alter their incentive responses as is done when paying on behalf of consumers for access to the good or service. Not surprisingly, we find that the second-best payment policies take on a very different form depending on the objectives of the firm and consumer access to the good. However, somewhat surprisingly, we find that the regulator can completely attenuate the firm's informational advantage when the good is nonmarketed and the firm is output-maximizing. The findings suggest that in those regulated markets that include a mix of for-profit and not-for-profit firms, the regulator must tailor its policies carefully with the firm's objectives if it is to achieve an optimal outcome. Offering the two types of firm the same contract, as is currently done with Medicare reimbursements for example,[6] is clearly sub-optimal.[7] The second-best outcome trades a reduction in distortions to output caused by the firm's information advantage, with the deadweight loss of 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 service and when it is not.

As has long been understood in Bayesian mechanism design problems of the type studied here, 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' best-response to the regulator's payment rules have not been thoroughly explored.[8] For example, the key finding of Lewis and Sappington (1988a) 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. 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) find that the firm is able to extract an information rent resulting in an output that is *always* distorted below 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 from the social optimum either. Caillaud et al. (1988) first note this with an example of hidden effort and unobserved firm cost. There is no distortion

[6]Strictly speaking, nonprofit and for-profit hospitals are reimbursed differently by Medicare in the sense that their unit payments are equivalent, but the tax-exemption status of nonprofit hospitals results in different fixed-payments.

[7]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.

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

away from the social optimum in this case because the agent internalizes all of the gains from exerting effort. Consequently, the agent 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. However, 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.[9]

In addition to having a topical relevance, we are also interested in analyzing contract design in the presence of unverifiable quality because it represents a dimension of moral hazard that interacts with the consumers. Unlike with unobservable effort, 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. This difference highlights the importance of accounting for the consumers' response, in addition to the firm's in the contract design. Another interesting characteristic of utilizing quality as the source of moral hazard is that the firm's optimal choice of quality is directly linked to the state parameter (i.e., the source of adverse selection), thus allowing us to study how the linkage between the two affect the contract.

The derivation of the paper's findings progresses as follows. In Section 2 the primary model is developed, including the derivation of the *quality-adjusted* cost and demand functions. Section 3 analyzes the optimal regulatory policies for a profit-maximizing firm and is divided into two parts. Section 3.1 analyzes the regulator's problem when it chooses to have consumers pay directly for the good or service, and Section 3.2 analyzes the regulator's problem when it pays for the good on behalf of consumers. Section 4 performs a similar analysis for an output-maximizing, nonprofit firm. Section 5 shows that the results are robust to demand uncertainty. Finally, in Section 6 we summarize the findings of the paper and discuss some applications and caveats.

Before proceeding, we distinguish this work from several related studies. As we utilize the standard solution techniques of Bayesian mechanism design the mechanics of the paper are the same or similar to those found in the previously cited papers. However, 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 the 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 nonprofit firms, nor do they fully characterize the distortions from the social-optimum.

The models of Baron and Myerson (1982) and Lewis and Sappington (1988a) are quite similar to one another and to the current model with respect to the information possessed by the regulator. In Baron and Myerson (1982) the firm has superior knowledge of its costs, and in Lewis and Sappington (1988a) the firm has superior knowledge of the demand. Neither model considers a dimension of moral hazard, quality or otherwise, and an important result of this paper is that we show that neither model is robust to the inclusion of quality. Lewis and Sappington (1988b) expand on Lewis and Sappington (1988a) by adding a second dimension of adverse selection: uncertainty in the firm's cost in addition to uncertainty in the market demand. Although only a single dimension of adverse selection is considered here, because quality affects demand, uncertainty in cost creates

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

uncertainty in demand and, likewise, uncertainty in demand creates uncertainty in cost allowing for consistency between to two sources of uncertainty. However, because the source of uncertainty is still just a single parameter, this remains a one-dimensional screening model. The results of Lewis and Sappington (1988b) would be similar to some of the results here if, in their model, the adverse selection parameters were perfectly correlated with one another.

Aguirre and Beitia (2004) modify the model of Lewis and Sappington (1988a) by making the consumer responsible for the unit payment, and the regulator responsible for the transfer payment. We also modify the source of payment to explore the role of consumer incentive responses to the regulator's payment policy by considering a marketed and nonmarketed good, but either the consumer or the regulator is responsible for both payments. 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.

Finally, Allen and Gertler (1991) analyze the welfare implications of fixed-price regulation for a marketed and nonmarketed good. There are two types of consumers, which correspond to the high and low demand states of Section 5 in the current model. Under a fixed unit payment rule the price is constant resulting in an under- and over-provision of quality for the high and low severity types respectively when the good is marketed. For a nonmarketed good, the firm provides an even lower level of quality since, at every quality level, demand is higher given the effective price of zero and the marginal gain to the firm of increasing quality is thus lower. We do not compare the level of quality between marketed and nonmarketed goods as our focus is on the degree of distortion from the social optimum, which itself varies with the status of the good. However, the difference in qualities should be in the direction of Allen and Gertler (1991) since the incentive response of the consumers and firm qualitatively the same across models.

2 The General Model

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 not verifiable so cannot be directly contracted upon.[10] We start by assuming that the firm has superior knowledge of its cost and later will show that the results are robust to asymmetric information about consumer demand. Consumer demand, $x(p, q)$, is a function of the price p and the quality, q, of the good.[11] [12]

[10] 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.

[11] 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. Moreover, in a setting such as for hospital services, the relevant observer of quality may be a health maintenance organization (HMO), which upon observing the quality of a hospital makes the decision of whether or not to add the hospital to its network, thus affecting the demand for the hospital's services. Supporting the notion that some attributes of quality are observed by HMOs, Gowrisankaran and Town (2003) present evidence that competition between hospitals increases the quality of care for HMO patients.

[12] 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.

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. When interpreting $x(p, q)$ as residual demand given all other firms supply some equilibrium level of quality, the regulator is assumed to take the number of firms in the market as given.[13] Demand is $\mathcal{C}^2$, increasing and concave in q,[14] and decreasing in p.

Gross consumer surplus is represented by the function $S(x, q)$.[15] S may reflect the consumers' direct value of consumption (i.e., $S(x, q; \theta) = \int_0^x P(\tilde{x}, q; \theta) d\tilde{x}$ where P is inverse demand) or S 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 surplus is increasing and concave in the quantity and quality; $S_x > 0$, $S_q > 0$, $S_{xx} \leq 0$, and $S_{qq} \leq 0$.[16] Moreover, marginal consumer surplus is weakly increasing in quality, $S_{xq} \geq 0$, reflecting the complimentarity of quality and output.

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 line. The cost of production is $\mathcal{C}^2$ and increasing and weakly convex in both quantity and quality and is increasing in θ. Moreover, the 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. Lastly, to insure the firm's problem is well-behaved, $c(x, q)$ is strictly quasi-concave.

The regulator is a Stackelberg leader and, as quality is an unverifiable attribute and the equilibrium quantity is not observable, is 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. The prices are enforceable by the regulator.[17]

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.[18]

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. Ensuring that the regulator's problem is well-behaved, F satisfies the monotone hazard rate condition, $d\{(1-F(\theta))/f(\theta)\}/d\theta < 0 < d\{F(\theta)/f(\theta)\}/d\theta$. The characteristics of the regulator's uncertainty are common knowledge.

2.1 Quality-Adjusted Cost and Value

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

$$(1) \qquad \max_q \{\Pi(p, T; \theta) = px(q, p) - c\big(x(q, p), q; \theta\big) + T\}.$$

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

[14]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.

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

[16]Subscripts represent partial derivatives.

[17]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.

[18]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.

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 properties of $x(q,p)$ are sufficient to insure the existence of $q(x,p)$.[19]

An intuitive way of viewing the firm's problem is that for a given price and demand state, it selects the level of quality that maximizes its profit. However, it is equivalent to view the firm 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

$$\max_x \{\Pi(p,T;\theta) = px - c\big(x, q(p,x);\theta\big) + T\}, \tag{2}$$

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

Similar to Rogerson (1994),[20] define $g(x;p,\theta)$ as the firm's *quality-adjusted* cost function

$$g(x;p,\theta) = c\big(x, q\big(x,p\big);\theta\big). \tag{3}$$

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. It will prove to be more intuitive to conduct the analysis using the quality-adjusted cost function instead of the primitive $c(\cdot)$. 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 reacting to the price response of the consumers:

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

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. However 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.

[19] 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 $q = f_2(x,p)$.

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

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

$$V(x;p) = S\big(x, q(x,p)\big). \tag{4}$$

As with the quality adjusted cost function, $V(x;p)$ denotes the consumer value of consuming the quantity x given that the quality has been adjusted to induce a demand of x when the unit price is p. Given the properties of S, it must be the case that V is increasing and concave in x.

3 Regulating a Profit-Maximizing Firm

3.1 Marketed Good

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

3.1.1 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 by deriving the optimal outcome when the regulator and firm have common knowledge of the cost or demand state, 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 or demand state,[21] and we refer to the solution as second-best when the firm has superior knowledge regarding the cost-state.

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

$$W(p,T;\theta) = \alpha CS(p,T) + (1-\alpha)\Pi(p,T;\theta). \tag{5}$$

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

$$CS = V(x;p) - (px + T). \tag{6}$$

By substituting Π and CS into (5) and rearranging, the socially optimal outcome is determined by the maximization program:[22]

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

[21] This is sometimes referred to as a constrained first-best since it is not truly a first-best solution given that the regulator cannot control consumer demand.

[22] 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.

where $\lambda = (2\alpha - 1)/\alpha > 0$. By taking the FOCs, the socially optimal outcome is defined as follows.[23]

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:*

$$V_x = g_x, \tag{7a}$$

$$V_p = g_p, \tag{7b}$$

$$\Pi = 0. \tag{7c}$$

3.1.2 Symmetric Information about θ

When the regulator cannot observe the 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 profit. The first order condition from the firm's problem, (2), shows that the firm's maximizer x^* solves

$$p = g_x\big(x^*(p, \theta); p, \theta\big). \tag{8}$$

As natural as this result is, it should be noted that with a marketed good, if the firm has no ability to adjust quality, then it cannot equate its (quality-adjusted) marginal cost with the unit price. In fact, without the ability or need to adjust quality, the firm has no decision to make as the price fully determines the quantity sold.[24] However, with the presence of quality, the firm is able to shift the demand and quality-adjusted cost curves to optimize profit.

The regulator's problem (RP-M) can be expressed as[25]

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

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

[24]For this reason, in models without a choice of quality, the unit payment will differ from marginal cost; e.g. Baron and Myerson (1982).

[25]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

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

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. The marginal benefit consists of the benefit derived from altering the equilibrium output, $V_x(dx^*/dp)$, and the benefit gained (or lost) due to the consumers' demand response to the change in unit price. Similarly, the marginal cost consists of the change in cost from altering the equilibrium output, $g_x(dx^*/dp)$, and the change in the cost level due to adjusting quality as a best-response to the consumers' demand response to the price change, g_p.

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 p and the 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 and demand information consists of the unique unit price $p^{fb}(\theta)$ and transfer payment $T^{fb}(\theta)$ satisfying:*

$$\begin{aligned} p^{fb}(\theta) &= g_x\big(x^{fb};p^{fb}(\theta),\theta\big) = V_x\big(x^{fb};p^{fb}\big) + \frac{V_p\big(x^{fb};p^{fb}(\theta)\big) - g_p\big(x^{fb};p^{fb}(\theta),\theta\big)}{dx^*\big(p^{fb}(\theta),\theta\big)/dp}, \\ T^{fb}(\theta) &= g\big(x^{fb};p^{fb}(\theta),\theta\big) - p^{fb}(\theta)x^{fb}, \end{aligned}$$

for all $\theta \in \Theta$.

The reason for the particular form of the unit payment is as follows. The two terms for $p^{fb}(\theta)$ account for the direct and indirect social costs and benefits to adjusting the 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 value of changing the equilibrium quantity. However, 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 social value of the quality adjustment, $V_p - g_p = (S_q - c_q)q_p$. As the firm is rewarded for each unit sold and not directly for the quality level of the good it must be compensated for each unit of additional surplus so the total change in social surplus due to the change in quality level is divided by the change in the equilibrium quantity, dx^*/dp.

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, then the first-best and socially optimal unit prices must also be equivalent. However, an important result similar to one first reported by Spence (1975) and Baron (1981) is that, because the firm chooses x so that $\Pi_x = 0$, that is, because the firm maximizes based on the marginal consumer's valuation of quality, the level of quality will necessarily be undersupplied. In the present model, quality may be oversupplied if there are negative externalities to consumption.[26] The following proposition formally reports this result.

[26] In Baron (1981) the regulator's value function is simply the area under the demand curve; i.e., $S = \int_0^x P(\tilde{x}, q)d\tilde{x}$ 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. However, if the regulator's measure of consumer value accounts

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

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

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

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 simultaneously control the firm's choice in quality while simultaneously adjusting consumer demand to achieve a socially optimal outcome.

3.1.3 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 an additional constraint on the firm's profit function: the single-crossing property (SCP).

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

Without loss of generality, we will 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 value function U. Because the firm's objective is to maximize profit, its value function is defined as

$$U(\hat{\theta}, \theta) = p(\hat{\theta}) x^*(p(\hat{\theta}), \theta) - g\big(x^*(p(\hat{\theta}), \theta); p(\hat{\theta}), \theta\big) + T(\hat{\theta}).$$

The regulator's objective is to maximize total expected social surplus subject to standard individual rationality and incentive compatibility constraints. The regulator's problem may be expressed as

$$\max_{p(\theta), U(\theta)} \int_\Theta \Big\{ V\big(x^*(p(\theta), \theta); p(\theta)\big) - g\big(x^*(p(\theta), \theta); p(\theta), \theta\big) - \lambda U(\theta) \Big\} dF(\theta)$$

subject to

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

---

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.

where $U(\theta) = U(\theta, \theta)$.

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*

(i) $\dfrac{dU(\theta)}{d\theta} = \dfrac{\partial U(\theta)}{\partial\theta} = -g_\theta\big(x^*(p(\theta), \theta); p(\theta), \theta\big)$,

(ii) $sign[dp\,/d\theta] = sign\left[\frac{\partial}{\partial\theta}\left(\Pi_p/\Pi_T\right)\right]$.

The proof for Lemma 1 is standard in the literature[27] so is placed in the appendix. Condition *(i)* is an application of the envelope theorem and states that the change in the firm's profit across cost-states must equal $-g_\theta = -c_\theta < 0$. The condition is a consequence of the fact that the firm will lower the service quality in higher cost-states unless it receives a higher unit payment. This gives the firm an incentive to misreport the state as being higher than it is to exaggerate its costs, thus incentive compatibility requires the firm receive higher rents in lower cost states. Condition *(ii)* is specific to the use of the two-part tariff and says that any payment policy satisfying incentive compatibility must offer a unit payment which 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. Notably, however, it is not restricted to being either an increasing or a decreasing function of the state parameter.

Condition *(i)* of Lemma 1 indicates that the firm's utility is decreasing in the state parameter, therefore using the integral form of the envelope theorem the firm's value function can be expressed as

$$U^*(\theta) = U^*(\overline{\theta}) + \int_\theta^{\overline{\theta}} g_\theta(x^*; p, \tilde{\theta})\ d\tilde{\theta}. \tag{10}$$

Because the firm extracts an information rent in all states but the highest, individual rationality binds in only the state $\overline{\theta}$ and the regulator can design the contact so that $U^*(\overline{\theta}) = 0$.[28]

By plugging (10) into the regulator's objective function and integrating by parts, the regulator's problem can be rewritten as

$$\max_{p(\theta)} \int_{\underline{\theta}}^{\overline{\theta}} \left\{V\big(x^*(p(\theta), \theta); p(\theta)\big) - g\big(x^*(p(\theta), \theta); p(\theta), \theta\big) - \lambda H(\theta) g_\theta\big(x^*(p(\theta), \theta); p(\theta), \theta\big)\right\}\ dF(\theta),$$

where $H(\theta) \equiv F(\theta)/f(\theta)$, and subject to condition *(ii)* of Lemma 1.

To solve the regulator's problem, we first ignore condition *(ii)* of Lemma 1 and verify that it is satisfied by the payment rule. The first-order condition of the regulator's optimization program yields

$$V_p\big(x^*; p(\theta)\big) - g_p\big(x^*; p(\theta), \theta\big) + \Big(V_x\big(x^*; p(\theta)\big) - g_x\big(x^*; p(\theta), \theta\big)\Big)\tfrac{dx^*}{dp} = \lambda H(\theta)\tfrac{d}{dp}\{g_\theta(x^*; p, \theta)\}. \tag{11}$$

The price p^{sb} solving Eq. (11) is the second-best price given the regulator's constraints. The interpretation of Eq. (11) is as follows. Increasing the payment to the firm in cost-states $[\theta, \theta + d\theta]$,

[27] See for example Guesnerie and Laffont (1984) and Caillaud et al. (1988)

[28] Or similarly the regulator could set a minimum profit level $\underline{U}$ for the firm and set $U^*(\overline{\theta}) = \underline{U}$.

which number $f(\theta)d\theta$, by dp increases the social surplus by $\big[V_p(x^*;p,\theta)-g_p(x^*;p,\theta)+\big(V_x(x^*;p,\theta)-g_x(x^*;p,\theta)\big)\frac{dx^*}{dp}\big]dp$. However, from *(i)* of Lemma 1, the increase in output simultaneously increases the firm's rent in cost-states $[\theta,\overline{\theta}]$, which number $F(\theta)$, by $\frac{d}{dp}\{g_\theta(x^*;p,\theta)\}dp$. The total social cost of the increase in the firm's rent is $\lambda F(\theta)\frac{d}{dp}\{g_\theta(x^*;p,\theta)\}dp$. Eq. (11) equates the marginal change in social surplus to the marginal social loss.

Recall that given p, a profit-maximizing firm will select the quantity equating the marginal cost to the unit price, $p = g_x\big(x^*(p,\theta);p,\theta\big)$. By substituting p for g_x in (11) and solving for p we can derive the second-best payment policy, but only if the resulting unit payment satisfies condition *(ii)* of Lemma 1. Otherwise the payment rule is not incentive compatible and the regulator will be unable to extract any usable information from the firm. Consequently, permitting the firm to use its private knowledge to select the unit price will be too socially costly. In such a case the regulator cannot allow the firm to choose the price and must set a fixed-price, $p(\theta) = \hat{p} \in [p^{sb}(\underline{\theta}), p^{sb}(\overline{\theta})]$.[29] Caillaud et al. (1988) and Laffont and Tirole (1993, pg. 161) refer to this as "the phenomenon of nonresponsiveness of the allocation with respect to private information."

To provide insight into the characteristics of the problem permitting the payment rule determined by Eq. (11) 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 profit due to increasing the price by dp can be decomposed into the marginal benefit x 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. However, more importantly, 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 2. *The p^{sb} solving (11) is incentive compatible only if*

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

Determining if the preferences of the firm and regulator for the direction of price adjustment are equivalent requires establishing functional forms for the cost, value, and demand functions or making several strict assumptions on the higher order derivatives of these functions.[30,31] However, to maintain generality and to avoid making assumptions on higher-order derivatives that do not

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

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

[31]See Rogerson (1987) for the analysis of when the preferences are equivalent for a similar principal-agent problem.

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 2 is satisfied.

Returning to the first-order condition of the regulator's problem given by Eq. (11) we can analyze how the firm's information rents distort the prices and output. Recall that the first-best solution requires the left-hand side of (11) to be set equal to zero. However, because the firm's information rent may be either increasing or decreasing with the unit payment, the right-hand side of (11) 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 shades the price up or down from the first-best price to limit the information rents attained by the firm. That is, if the firm's rents are increasing with the unit payment ($d\{g_\theta\}/dp > 0$), 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:*

$$p^{sb} \begin{cases} < p^{fb} & when \quad d\{g_\theta(x^*; p, \theta)\}/dp > 0, \\ = p^{fb} & when \quad d\{g_\theta(x^*; p, \theta)\}/dp = 0, \\ > p^{fb} & when \quad d\{g_\theta(x^*; p, \theta)\}/dp < 0. \end{cases}$$

Proof. Because the regulator's problem is quasiconcave in x and p (see footnote 25) 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. (11) 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 $p^{sb} > p^{fb}$. Therefore, when $d\{g_\theta(x^*; p, \theta)\}/dp < 0$ we have $p^{sb} > p^{fb}$ and when $d\{g_\theta(x^*; p, \theta)\}/dp > 0$ we have $p^{sb} < p^{fb}$. □

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 (ergo the quality) may be over- or undersupplied relative to the first-best as well. To see how the second-best output compares to the first-best, recall that by integrating by parts we have the identity $\int_{\underline{\theta}}^{\overline{\theta}} U(\theta)dF(\theta) = \int_{\underline{\theta}}^{\overline{\theta}} H(\theta)g_\theta(x^*; p, \theta)dF(\theta)$, which implies $sign[dU(\theta)/dp] = sign[d\{g_\theta(x^*; p, \theta)\}/dp]$. Therefore to identify how the firm's rents change with price we can decompose $d\{g_\theta\}/dp$ into its two constituent parts:

$$\frac{d}{dp}\left\{\frac{\partial g}{\partial \theta}\right\} = \frac{\partial^2 g}{\partial \theta \partial p} + \frac{\partial^2 g}{\partial \theta \partial x}\frac{dx^*}{dp}.$$

The first term, $g_{\theta p}$, identifies the direct change in the firm's rent that follows from an increase in the unit price. The partial change in the firm's rents following a price change (in all cost states) is the change in the firm's rent that comes about from adjusting quality to maintain the *same* equilibrium output. Increasing quality increases the cost of production therefore the firm's rents 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 therefore 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 must be undersupplied relative to first-best because the firm's rents are unambiguously increasing with the unit payment. In consequence, the regulator will set a price below first-best 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. However, 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.

Proposition 4. *With cost uncertainty, the relative size of the second- to first-best price and output is determined by the rules:*

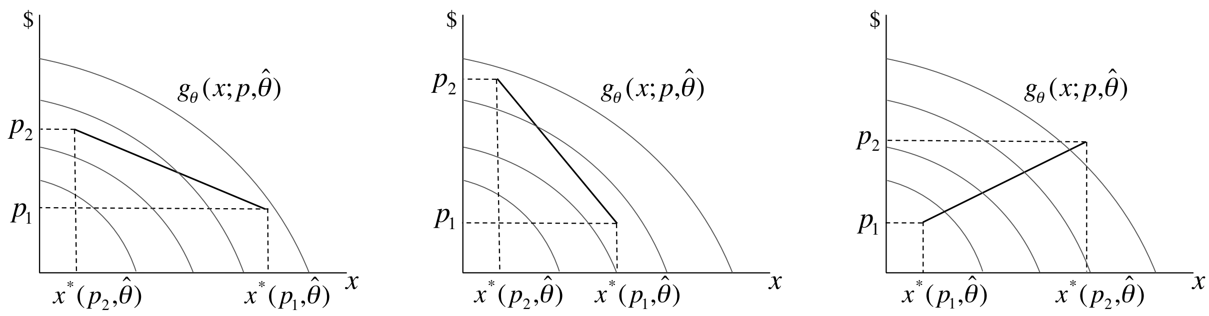
$$
\begin{array}{rlll}
\textit{(i)} & dx^* /dp < -(g_{\theta p}/g_{\theta x}) < 0 & \Rightarrow d\{g_\theta\} /dp < 0 \Rightarrow p^{sb}(\theta) > p^{fb}(\theta) & \Rightarrow x^{sb}(\theta) < x^{fb}(\theta), \\
\textit{(ii)} & -(g_{\theta p}/g_{\theta x}) < dx^* /dp < 0 & \Rightarrow d\{g_\theta\} /dp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

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

Proof. The proof follows immediately from Proposition 3 and from identifying $sign[d\{g_\theta\} /dp]$. □

Figure 1 provides a graphical representation of 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 $g_\theta(x^*; p, \hat{\theta})$ are displayed and can be thought of as iso-rent curves. 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) represent case *(i)* of proposition 4 as the decrease in output is sufficient to drop to a lower level curve resulting in an decrease in the firm's rents. In figure 1(b) the decrease in output is insufficient. Representing case *(ii)*, the decrease in output results in a jump up to a higher level curve and an increase in rents. 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, in some states the regulator's problem will satisfy condition *(i)* and for others condition *(ii)* of Proposition 4.

The importance of Proposition 4 is that it tells us that we cannot predict *a priori*, without having specific functional forms for the cost, demand, and value functions, how the firm's informational advantage will distort output when quality is added to the model. Moreover, the qualitative result is decidedly different than that found in Baron and Myerson (1982) and others where the informational advantage results in a strictly downward distortion. With its choice of quality, the firm is able to adjust the quantity demanded for a given unit price. In consequence, the unit price



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

Figure 1: Graphical Depiction of Proposition 4

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. This compound adjustment in demand may result in either an over- or under-supply relative to first best.

3.2 A Nonmarketed Good

When the good or service is not marketed then there is no direct demand-response to the contract and the regulator need only account for how the firm best-responds to the payment rule it sets. In this section we show that this simplifies the regulator's problem generating a uni-directional distortion in the outcome, consistent with the extant literature on regulation under asymmetric information.

3.2.1 Symmetric Information about θ

When the regulator pays for the good directly, its problem continues to be that of designing a menu of two-part tariffs $\{p(\theta), T(\theta)\}_{\theta\in\Theta}$ 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:

$$CS = S(x,q) - (1+\gamma)(px + T).$$

As with a marketed good, it is convenient to define and work with the quality-adjusted cost and consumer value functions. However, because consumer demand is unaffected by the unit price, the quality-adjusted functions are parameterized by only the state parameter; i.e.,

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

Again, $g(x;\theta)$ represents the cost of producing x in state θ and $V(x)$ represents the consumer value of x given the quality is adjusted to insure an equilibrium demand of quantity x.

By substituting Π and CS into (5) and rearranging, the regulator's problem (RP-NM) when knowledge of the cost is symmetric can be expressed as

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

Of course 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. 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.[32] This follows 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 5. *The optimal payment rule for a nonmarketed good with symmetric cost and demand information consists of the unique unit price $p^{fb}_{nm}(\theta)$ and transfer payment $T^{fb}_{nm}(\theta)$ satisfying:*

$$p^{fb}_{nm}(\theta) = g_x\big(x^*\big(p^{fb}_{nm}(\theta),\theta\big),\theta\big) = \tfrac{1}{1+\gamma}V_x\big(x^{fb}\big),$$
$$T^{fb}_{nm}(\theta) = g\big(x^{fb},\theta\big) - p^{fb}_{nm}(\theta)x^{fb},$$

for all $\theta \in \Theta$.

Because price is not present in the demand function, the first-best 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 alone, 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.[33]

3.2.2 Asymmetric Information about θ

When information is asymmetric, the regulator's problem is defined as

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

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

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

[33] 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.

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 3.

Lemma 3. *When the good is nonmarketed, the SCP is satisfied and* $d\{\Pi_p/\Pi_T\}/d\theta < 0$ *for all* $\theta \in \Theta$.

Proof. Taking derivatives of the firm's profit with respect to price and transfer payment gives $\Pi_p/\Pi_T = x^*$. From the conjugate pairs theorem $sign[dx^*/d\theta] = sign[\Pi_{x\theta}]$ and $\Pi_{x\theta} = -g_{x\theta} < 0$. □

Any incentive compatible mechanism must still satisfy Lemma 1.[34] However, 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.

Lemma 4. *When the good is nonmarketed, the payment policy* $\{p(\theta), T(\theta)\}_{\theta\in\Theta}$ *is incentive compatible only if* $dp/d\theta \leq 0$.

Proof. See appendix. □

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.

From condition *(i)* of Lemma 1 the firm's utility continues to be increasing in the cost state when the good is nonmarketed and the firm's value function can be expressed as Eq. (10). By plugging (10) into the regulator's objective function (and integrating by parts), the regulator's problem can be rewritten as

$$\max_{p(\theta)} \int_{\underline{\theta}}^{\overline{\theta}} \left\{ V\big(x^*(p(\theta),\theta),\theta\big) - (1+\lambda)g\big(x^*(p(\theta),\theta),\theta\big) - (\lambda+\gamma)H(\theta)g_\theta(x^*;\theta) \right\} dF(\theta),$$

subject to $dp/d\theta \leq 0$.

To solve the regulator's problem, we first ignore the constraint $dp/d\theta \geq 0$ and verify that it holds at the optimum. The first-order condition of the regulator's optimization program yields

$$V_x\big(x^*\big(p(\theta),\theta\big),\theta\big) - (1+\gamma)g_x\big(x^*\big(p(\theta),\theta\big),\theta\big) = (\lambda+\gamma)H(\theta)g_{\theta x}(x^*;\theta). \quad (12)$$

The quantity x^* solving Eq. (12) is the second-best quantity given the regulator's constraints. The interpretation of Eq. (12) is similar to the interpretation of Eq. (11), the first-order condition for a marketed good. The main difference between the two FOCs is that, 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 nonmarketed good. The lack of demand response implies $g_{\theta p} = 0$ and because $dx^*/dp > 0$, the characteristics of the problem are reduced to rule *(iii)* of Proposition 4. Graphically the problem is represented by figure 1(c) where the level curves are transformed into vertical lines. This result is stated in the following proposition.

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

Proposition 6. *When the good is nonmarketed and the payment rule is not constant, 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.*

Because the equilibrium quantity demanded is uniquely determined by the quality chosen by the firm, an equilibrium output below first-best necessarily requires that quality is set below first-best as well. Therefore, the optimal regulation of a firm via a nonmarketed good results in unambiguously lower quality.

Finally, it is worthwhile to note that the equilibrium output is equal to the first-best output when the firm's quality-adjusted cost function exhibits constant differences across cost states; i.e., when $g_{x\theta}(x, \theta) = 0$. The firm's cost function will exhibit constant differences whenever the cost parameter affects only the fixed cost of production; i.e., when the cost function can be represented as additively separate: $c(x, q) = c_1(x, q) + c_2(\theta)$. When $g_{\theta x} = 0$ the quality-adjusted marginal cost curve is the same for all values of the state parameter and the regulator knows precisely what level of quality the firm will set for all cost states. Nevertheless, the regulator must still insure incentive compatibility by leaving the firm with an information rent via the lump-sum transfer.

4 A Nonprofit Firm

The purpose of this section is to explore how the firm's market conduct alters the optimal contract, and, more importantly, what outcome the regulator can achieve. The objective of a nonprofit firm varies with the market within which the firm produces. However, we limit our attention to health care markets and assume,[35] à la Newhouse (1970),[36] that the firm's objective is to maximize gross consumer surplus. This does not mean, however, that the incentives of the firm and regulator are in lock-step with one another. Because the nonprofit firm's objective is to maximize the gross consumer surplus whereas the regulator's objective is to maximize net social surplus, there exists a misalignment in their objectives and the firm prefers to provide a higher quality of the good than what the regulator considers socially optimal. As a Stackelberg leader, the regulator must still design the payment policy to provide the correct incentives based on the firm's objectives.

4.1 Symmetric Information about θ

To see how the firm's objective alters what the regulator can achieve, first consider the firm's allocation rule. Based on a payment rule $\{p, T\}$, the firm will choose the quantity $x^*(p, T, \theta)$ solving

$$px^* - g(x^*; p, \theta) + T = 0.$$

[35] In 2004 the health care sector accounted for 58.7 percent of all nonprofit revenue. The next largest sector is education, accounting for 16.3 percent of nonprofit revenue (National Center for Charitable Statistics, 2007).

[36] Newhouse (1970) argues that a nonprofit hospital's objective is prestige which comes through the quantity and quality of the service provided. In the present model that translates into output maximization. Needleman (2001) provides a concise survey of research exploring what other objectives a nonprofit hospital may have. For example Davis (1972) postulates an objective of cash flow maximization, Pauly and Redisch (1973) postulate the nonprofit's objective is to promote the welfare of the medical staff, Rose-Ackerman (1987) postulates an objective of meeting donor expectations, and Frank and Salkever (1991) postulate an objective of reducing unmet need in the community.

In other words, the firm will continue to increase output until it breaks even.[37] Thus, the firm's optimizer, x^*, is a function of both price instruments. In the case of a marketed good, consumers' respond to only the unit price permitting the regulator the use of the transfer payment to achieve the socially optimal outcome. It does so by using the unit price to control consumers by setting it to the socially optimal level, $p^{fb} = p^{so}$ and using the transfer payment to induce the firm to produce the level of quality that induce a demand of x^{so} at p^{so}. Of course, if the regulator can achieve the socially optimal outcome for a marketed good, then it can do so with a nonmarketed good as well due to the lack of demand response. In fact, as Proposition 7 reports, for a nonmarketed good the optimal payment policy is not unique as the regulator can set any unit and transfer payment satisfying $px + T = g(x^{so};\theta)$ as long as the unit payment is less than or equal to the marginal cost of production at x^{so}.

Proposition 7. *The optimal regulatory policy under symmetric cost and demand information consists of the unique unit price $p^{fb}_{np}(\theta)$ and transfer payment $T^{fb}_{np}(\theta)$ satisfying:*

(i) Marketed Good:

$$\begin{aligned} p^{fb}_{np}(\theta) &= p^{so}(\theta), \\ T^{fb}_{np}(\theta) &= g(x^{so}; p^{so}, \theta) - p^{so}x^{so}. \end{aligned}$$

(ii) Nonmarketed good:

$$\{p^{fb}_{np}(\theta), T^{fb}_{np}(\theta)\} \in \big\{\{p, T\} \mid 0 \le p \le g_x(x^{so}, \theta) \text{ and } T = g(x^{so}, \theta) - px^{so}\big\}.$$

Proof. See appendix. □

4.2 Asymmetric Information about θ

When there is adverse selection the nonprofit firm has an incentive to misreport the demand state, even though it does not care about profit, if doing so allows it to induce a higher equilibrium output. Thus, any payment policy must continue to satisfy incentive compatibility. Because the firm chooses the quantity leaving it with zero profit, an incentive compatible policy is the policy, with which the firm's output is maximal when it announces the true demand state.

Definition 3. *A menu of two-part tariffs $\{p(\theta), T(\theta)\}_{\theta\in\Theta}$ is incentive compatible for an output-maximizing firm when the following holds for all $\hat{\theta}, \theta \in \Theta$ where θ is the true state and $\hat{\theta}$ is the firm's announcement:*

$$x^*\big(p(\hat{\theta}), T(\hat{\theta}), \theta\big) \le x^*\big(p(\theta), T(\theta), \theta\big).$$

Utilizing the definition of incentive compatibility for an output-maximizing firm, the following lemma characterizes the necessary and sufficient conditions for an incentive compatible payment policy for a nonprofit firm.

Lemma 5. *The menu of two-part tariffs $\{p(\theta), T(\theta)\}_{\theta\in\Theta}$ is incentive compatible if and only if it satisfies the conditions:*

[37] Instead of a break-even constraint, the firm could want some minimum level of profit $\underline{\pi}$ to use for other projects; e.g., a hospital may want to be able to provide a certain level of charitable care. Including a minimum profit level does not alter the analysis.

(i) $\dfrac{dx^*}{d\theta} = \dfrac{\partial x^*}{\partial \theta} = -\dfrac{g_\theta(x^*; p, \theta)}{g_x(x^*; p, \theta) - p} < 0,$

(ii) $sign[dp\,/d\theta] = sign\big[\frac{\partial}{\partial\theta}\big\{(dx^*\,/dp)/(dx^*\,/dT)\big\}\big].$

Proof. See appendix. □

Lemma 5 is the nonprofit analog to Lemma 1 and the proof follows similarly. Condition *(i)* of Lemma 5 is an application of the envelope theorem and identifies the path the equilibrium quantity follows as the state parameter increases. Similar to Lemma 1, condition *(i)* identifies the information "rents" that the firm can extract due to it information advantage. However, with an output-maximizing firm these rents translate into output instead of profit. The interesting characteristic of condition *(i)* is that the information rents, which are defined by the path the firm's objective function x^* follows across cost states, is a direct function of the unit price established by the regulator. The regulator can take advantage of this by setting a unit price (and transfer payment) inducing the output path to follow the socially optimal path.

To see how the regulator might accomplish this, we start by identifying the first-best optimal path. Because the first-best output is the x^{fb} solving (RP-M) for a marketed good and (RP-NM) for a nonmarketed good, the first-best optimal path for output across demand states is defined as

$$\frac{dx^{fb}}{d\theta} = \begin{cases} \dfrac{g_{p\theta} - (V_{x\theta} - g_{x\theta})\frac{dx^{fb}}{dp} - (V_x - g_x)\frac{d^2x^{fb}}{dpd\theta}}{V_{px} - g_{px} + (V_{xx} - g_{xx})\frac{dx^{fb}}{dp}} & \text{(marketed good)} \\ \dfrac{(1+\gamma)g_{x\theta}}{V_{xx} - (1+\gamma)g_{xx}} & \text{(nonmarketed good)} \end{cases}$$

When the payment rule implements truthful revelation, the firm's choice of x can be expressed as

$$x^*(\theta) = x^*(\underline{\theta}) - \int_{\theta}^{\overline{\theta}} \frac{\partial x^*}{\partial \theta}(\tilde{\theta})d\tilde{\theta}.$$

Therefore, by choosing a payment policy which induces the firm to choose $x^*(p(\underline{\theta}), T(\underline{\theta}), \underline{\theta}) = x^{fb}(\underline{\theta})$ and equates the path of x^* with the first-best optimal path across demands states, the regulator will be able to induce the firm to choose the first-best output $x^{fb}(\theta)$ in every demand state $\theta \in \Theta$. Proposition 8 identifies the payment rule inducing the first-best optimal output for a nonprofit firm.

Proposition 8. *The menu of two-part tariffs* $\{p^{np}(\theta), T^{np}(\theta)\}_{\theta\in\Theta}$ *induces a not-for-profit firm to produce the socially optimal output* x^{fb}*, where*

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

and $sign[dp^{np}\,/d\theta] = sign[dx^{fb}\,/d\theta]$ *for all* $\theta \in \theta$*.*

Proof. See appendix. □

The payment rule reported by Proposition 8 is limited, for although the rule is sufficient to induce the first-best level of output, it does not generically achieve the first-best outcome as defined by the solution to Eq. (9). This is because the equilibrium quantity is determined by the price consumers must pay, and the level of quality the firm sets. 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. Essentially the firm has one instrument, with which to solve two equations and it cannot. The result is that for a given $x^{fb}(\theta)$, if $p^{np} > p^{fb}$ then quality is also set above the first-best level, and vice-versa.

On the other hand, when the good is nonmarketed, the level of quality uniquely identifies the equilibrium output x^*. As a result, the first-best level of output necessarily implies the first-best level of quality, Moreover, because there is no demand response to the unit payment, the first-best quantity is equivalent to the social optimum regardless of the firm's objective. Thus, as long as $dp^{np}/d\theta \geq 0$, then the payment rule reported by Proposition 8 will induce the socially optimal outcome when the firm is output-maximizing. The following proposition summarizes the previous discussion.

Proposition 9. *The regulator can induce the socially-optimal outcome if the firm is output-maximizing and the good is nonmarketed, otherwise the second-best outcome is distorted away from first-best.*

Proof. See appendix. □

The main finding reported by Proposition 9 is that, without gaining any additional information, the regulator can completely attenuate the firm's information advantage and induce it to produce in the social interest by altering how the contract interacts with the incentives of both the consumers and the firm. On the consumer side, the regulator disconnects the consumers from the contract by paying for the good with public funds and the firm's output is controlled because its objective makes output responsive to both payment instruments.

5 Asymmetric Information about Demand

As was discussed in the introduction, Lewis and Sappington (1992) find 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 Therefore it is worth exploring whether or not the present results are robust to the source of asymmetric information.

To see how demand uncertainty by the regulator alters the results we make the following changes to the model. First, we add the state parameter as an argument to demand, $x(q, p; \theta)$. To maintain consistency with the model under cost uncertainty, we assume that higher states result in *less* demand at the same service quality and unit price: $x_\theta < 0$. As before we will work with the inverse 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.

As before it will be more convenient to work with quality-adjusted cost and social value functions. The quality-adjusted cost function is now defined as:

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

The quality-adjusted marginal cost again takes into account the change in cost associated with increasing the quality to induce higher demand. 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,

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

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 states.

Starting with the profit-maximizing firm, condition *(i)* of Lemma 1 identifies a necessary condition for an incentive compatible payment rule. The condition, $dU/d\theta = -g_\theta(x; p, \theta) < 0$, applies regardless of the source of asymmetric information. Moreover, condition *(ii)* identifies a sufficient condition, which is dependent on the properties of the firm's profit. Because the properties of g do not change based on the source of information asymmetry it also applies to both cases. In consequence, the firm earns no rents in the high state $\overline{\theta}$ and there is no output distortion in the low state $\underline{\theta}$ regardless of the source of information asymmetry.

For a marketed good, the conditions that result in a second-best unit price that is greater or less than the first-best, as enumerated by Proposition 4, are independent of the source of asymmetric information. However, the relationship between prices is dependent on how the unit payment affects the firm's information rents. That is, the relationship is dependent on the sign for $d\{g_\theta\}/dp$, which does vary with the source of asymmetric information. To see this decompose $d\{g_\theta\}/dp$ into its component parts again:

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

Regardless of the source of information asymmetry $g_{\theta x} > 0$ and dx^*/dp is free. 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 demand uncertainty $g_{\theta p} = c_q q_{\theta p}$ and $sign[q_{\theta p}]$ is free. 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. However, 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 uncertainty 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., $g_{\theta p}/g_{\theta x} > 0$. However, with uncertainty in demand 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 demand uncertainty, 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 undersupplied 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. 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. 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.

Finally, because g_θ is independent of the source of information asymmetry, all of the results remain unchanged when the firm is output-maximizing. Reflecting this, the proofs include both cases of cost and demand uncertainty.

6 Discussion and Concluding Remarks

We have examined the optimal payment policies for a monopolist providing a good or service where demand is partially driven by 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. Moreover, to further complicate the regulator's problem, we have assumed that the firm possesses superior information regarding some aspect of the market, be it cost or demand. Within the same information environment the regulator can achieve strikingly different outcomes based on the consumers' access to the good and the firm's market conduct. Indeed, when the good is nonmarketed and the firm is not-for-profit the regulator can completely attenuate the information advantage of the firm. However, because there is a deadweight loss associated with using public funds, attenuating the firm's information advantage has a social cost and does not represent a panacea for the regulator. When the good is marketed the firm's output may be under- or oversupplied relative to the social optimum depending on the characteristics of the consumer value, demand, and cost functions.

The ambiguous direction of the distortion that occurs with a marketed good is a result of the fact that the firm can manipulate demand by adjusting quality. If the regulator raises the unit price then the firm adjusts the quality to compensate for the negative demand response, and to re-optimize its choice of output. The result is the regulator may have to discipline the firm by setting the price above first-best levels to make it more expensive to produce output, or set a price below first-best to decrease the unit revenue. In contrast, when the firm cannot manipulate demand by adjusting quality (because it is either not in the model, or consumer demand is inelastic to quality), then the unit price uniquely determines the quantity demanded. The distortion from the social optimum caused by the firm;s information advantage is always downward 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 in procurement models where demand is price-inelastic as the procurer seeks a unit of the good since higher payment always increases the firm's profit.[38]

The results of the model represent only the best possible outcomes and some caveats apply of course. Given the complex interactions of the cost, value, and demand functions, for many classes of functions, the optimal payment policies for the various scenarios considered may not

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

satisfy incentive compatibility. When they do not, the regulator will not be able to extract any information from the firm and is better off setting a constant payment rule, resulting in strong distortions away from the social optimum regardless of the type of good and objective of the firm. The optimal payment policy need not exhibit complete separation or complete pooling either, but instead there may be pooling for some subset of cost or demand states and separation for others. We have provided some insight into when the regulator cannot achieve type separation, however, only when the functional forms are known can we identify if the regulator can achieve the second-best outcome.

When drawing policy implications from the results of the model one should exercise some caution. Although it appears that the regulator can achieve the social optimum by restricting the firm to being not-for-profit, the result critically depends on the firm's underlying objective, which will vary across industries. For example, nonprofit hospitals follow a long tradition of religious affiliation and a desire to provide charity and emergency care.[39] Furthermore the Internal Revenue Service (IRS) monitors hospitals to insure they comply with "community benefit" standards; i.e., that the hospital acts in a way consistent with optimizing community value.[40] Similar institutional arrangements and traditions motivate nonprofit charter and private schools. However, it is unlikely that preventing a firm from earning and distributing profit to share holders will result in an altruistic objective for the firm if there is no institutional tradition of altruism in that market. As Oleck and Stewart (1994) argue, nonprofit status is appropriate only when altruistic, ethical, moral, or social motives are the clearly dominant objectives for the firm.

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, however, 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 Patient Protection and Affordable Care Act the government has mandated insurance coverage. To help those for whom premiums would exceed a certain percentage of income, the government provides subsidies, and Medicare Advantage can also be thought of as a voucher program, softening the price elasticity of demand for those eligible. Given the prevalence of such mixed payment systems, studying the optimal multitiered payment policy is an interesting and important avenue of future research.

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.

[39] In the early 20th century hospitals began as purely charitable organizations providing care to those who could not afford personal, home care resulting in their original tax-exempt status (Owens, 2005).

[40] The "community benefit" standard is a controversial measure of social benefit that allows hospitals to maintain their tax-exempt status without necessarily providing charity care. See Colombo (2005) for a history of the laws determining tax-exemption for hospitals and their flaws.

Appendix

PROOF OF LEMMA 1

Condition *(i)* is a necessary condition for an optimum as is derived as follows.

Recall the firm's utility function is defined as

$$U(\hat{\theta},\theta) = p(\hat{\theta})x^*\big(p(\hat{\theta}),\theta\big) - g\big(x^*\big(p(\hat{\theta}),\theta\big),\theta\big) + T(\hat{\theta}). \tag{A-1}$$

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

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

Because $\frac{\partial \Pi}{\partial x^*}(x^*;p,T,\theta) = p(\hat{\theta}) - \frac{\partial g}{\partial x}(x^*;p,\theta) \equiv 0$ by the envelope theorem, we have

$$\left.\frac{\partial U}{\partial \hat{\theta}}\right|_{\hat{\theta}=\theta} = \frac{dp}{d\theta}x^*\big(p(\theta),\theta\big) + \frac{dT}{d\theta} = 0. \tag{A-2}$$

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

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

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

$$\frac{dU}{d\theta} = -\frac{\partial g}{\partial \theta}. \tag{A-3}$$

Next, condition *(ii)* of the lemma represents a sufficient condition. To show sufficiency, start with

$$\frac{\partial U(\hat{\theta},\theta)}{\partial \hat{\theta}} = \frac{dp}{d\hat{\theta}}x - \frac{\partial g}{\partial p}\frac{\partial p}{\partial \hat{\theta}} + \frac{dT}{d\hat{\theta}}. \tag{A-4}$$

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

$$\frac{dT}{d\hat{\theta}} = \frac{\partial g}{\partial p}\frac{\partial p}{\partial \hat{\theta}} - \frac{dp}{d\hat{\theta}}x\big(p(\hat{\theta}),\hat{\theta}\big). \tag{A-5}$$

Plugging (A-5) into (A-4) yields

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

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

$$\text{(A-6)} \qquad \frac{\partial U(\hat{\theta}, \theta)}{\partial \hat{\theta}} = \frac{dp}{d\hat{\theta}} \frac{\partial^2 U\big(\hat{\theta}, \tilde{\theta}\big)}{\partial p \partial \tilde{\theta}} (\theta - \hat{\theta}).$$

Because $\Pi_T = 1$, the second-order cross partial derivative of Eq. (A-6) is equal to $\frac{\partial}{\partial \theta}(\Pi_p / \Pi_T)$. The condition $sign[dp\,/d\theta] = sign\left[\frac{\partial}{\partial \theta}(\Pi_p / \Pi_T)\right]$ implies

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

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

PROOF OF LEMMA 4

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 $sign[dp\,/d\theta] = sign\left[\frac{\partial}{\partial \theta}(\Pi_p / \Pi_T)\right]$ requires $sign[dp\,/d\theta] < 0$.

PROOF OF PROPOSITION 7

The nonprofit firm's objective is to maximize social value (by maximizing total output), subject to earning non-negative profits. The firm's optimization program is defined as

$$\text{(A-7)} \qquad \max_{x>0} V\big(x, \theta\big) \text{ subject to } \Pi(x; p, T, \theta) \geq 0.$$

Letting γ denote the Lagrange multiplier, the first-order conditions for the firm's problem yield

$$\text{(A-8a)} \qquad \gamma = \frac{V_x(x, \theta)}{g_x(x, \theta) - p},$$

$$\text{(A-8b)} \qquad px - g(x, \theta) + T = 0,$$

$$\text{(A-8c)} \qquad 0 < \gamma \leq \infty.$$

Because x must take a positive value, corner solutions are ignored. $\gamma > 0$ follows from the fact that $V_x > 0$ and $p \leq g_x(x^*, \theta)$, where x^* is the maximand of (A-7). This is true because if, to the contrary, p exceeds the quality-adjusted marginal cost of producing the quantity x^*, then the firm could induce a higher equilibrium quantity by increasing the quality of the good, all without losing profit. The Lagrange multiplier, γ, identifies the shadow price of increasing firm profit in terms of lost consumer benefit. The shadow price is decreasing in output ($\gamma_\theta < 0$) since consumers exhibit decreasing returns to quantity (and quality). Consequently, the further along the consumers' value function the firm is, the lower the cost to sacrificing consumer value for firm profits. Finally, when price equals the quality-adjusted marginal cost ($p = g_x$), the Lagrange multiplier will assume the value ∞.

Throughout the analysis we have used the quality-adjusted cost function $g(x, \theta) = c(x, q)$. 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

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

Because the quality-adjusted marginal cost is convex in x it is easy to show the following lemma.

Lemma 6. *There exists a unique $x \geq 0$ such that $AC_{qa}(x) = MC_{qa}(x)$.*

Let x^E denote the unique x satisfying $AC_{qa}(x) = MC_{qa}(x)$ and let $\hat{x} = \arg\max_x V(x,\theta) + \gamma\Pi(x,\theta)$, then the contract inducing the first-best outcome can be characterized by the relative value of x^* to x^E.

The following lemmas characterize the first-best policy.

Lemma 7. *A policy inducing the first-best must include a positive lump-sum transfer for all quantities $0 < x^* < x^E$.*

When the first-best outcome is less than the efficient scale then average costs exceed marginal costs. If the payment rule does not include a positive transfer then the unit price must exceed the average cost in order for the firm to produce any quantity. However, when the unit price exceeds the marginal cost, then the firm will continue to produce until at least the marginal cost equals the unit price. A positive transfer lowers the firm's average cost curve sliding the efficient scale down the marginal cost curve.

Lemma 8. *First-best can be induced with only a unit payment, p, when $x^E \leq x^*$.*

Lemma 8 corresponds with lemma 3.6 in Rogerson (1994) and follows from the fact that, when the first-best quantity is greater than the efficient scale, then the marginal cost exceeds the average cost at x^* and the regulator can induce the first-best by setting $p = AC_{qa}(x^*)$.

Lemma 9. *First-best can be induced with only a lump-sum transfer, T, for any $x^* > 0$.*

Lemma 9 follows immediately from (A-8b). 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.

Combining Lemmas 7 - 9 yields the pricing rule of the proposition.

PROOF OF LEMMA 5 There are two parts to this lemma. Starting with condition *(i)*, a necessary condition for optimization, define $x^*(\hat{\theta},\theta)$ as the x which maximizes $p(\hat{\theta})x - g(x;p(\hat{\theta}),\theta) + T(\theta)$. The firm's announcement is a function of the true state; i.e., $\hat{\theta} : \Theta \to \Theta$. Therefore, using the implicit function theorem, $\frac{dx^*}{d\theta}$ can be expressed as

$$\text{(A-9)} \qquad \frac{dx^*(\hat{\theta}(\theta),\theta)}{d\theta} = -\frac{p_\theta \frac{d\hat{\theta}}{d\theta} x - g_p p_\theta \frac{d\hat{\theta}}{d\theta} - g_\theta + T_\theta \frac{d\hat{\theta}}{d\theta}}{p - g_x} = \frac{dx^*}{d\hat{\theta}} \frac{d\hat{\theta}}{d\theta} - \frac{-g_\theta}{p - g_x}.$$

The firm announces the $\hat{\theta}$ which mazimizes output, therefore by the envelope theorem the first term on the RHS of (A-9) is zero. Furthermore, the proof for Proposition 7 establishes that $p < g_x$ and condition *(i)* of the lemma is satisfied.

To prove condition *(ii)* we start with the definition of incentive compatibility for an output-maximizing firm. From the definition, it must be the case that for any θ_1 and θ_2 in Θ where $\theta_1 < \theta_2$, the following hold

$$x^*\big(p(\theta_2), T(\theta_2), \theta_1\big) \leq x^*\big(p(\theta_1), T(\theta_1), \theta_1\big), \tag{A-10}$$

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

Adding (A-10) and (A-11) 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

$$\int_{\theta_1}^{\theta_2} \int_{\theta_1}^{\theta_2} \frac{d^2 x^*}{d\hat{\theta} d\theta} d\hat{\theta} d\theta \geq 0. \tag{A-12}$$

Because (A-12) 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

$$\frac{d^2 x^*}{dT d\theta} \frac{dT}{d\hat{\theta}} + \frac{d^2 x^*}{dp d\theta} \frac{dp}{d\hat{\theta}} \geq 0. \tag{A-13}$$

We can simplify (A-13) by observing that a truthful announcement of the state parameter is optimal if

$$\left.\frac{dx^*}{d\hat{\theta}}\right|_{\hat{\theta}=\theta} = \left.\frac{dx^*}{dp} \frac{dp}{d\hat{\theta}}\right|_{\hat{\theta}=\theta} + \left.\frac{dx^*}{dT} \frac{dT}{d\hat{\theta}}\right|_{\hat{\theta}=\theta} = 0. \tag{A-14}$$

Using (A-14) we can rewrite (A-13) as

$$\left.\frac{\partial}{\partial\theta} \left(\frac{dx^* / dp}{dx^* / dT} \right) \frac{dp}{d\hat{\theta}}\right|_{\hat{\theta}=\theta} \geq 0.^{41} \tag{A-15}$$

Eq. (A-15) is a special case of the condition derived in Theorem 1 of Guesnerie and Laffont (1984). The term $(dx^* / dp)/(dx^* / dT)$ is the nonprofit firm's MRS of unit payment for fixed transfer so the firm's objective function satisfies the SCP when $d\{(dx^* / dp)/(dx^* / dT)\} / d\theta$ is monotonic for all $\theta \in \Theta$. It is clear that when the firm's value function satisfies the SCP, then the payment policy is incentive compatible only if the change in the unit payment with the state parameter is of the same sign. Moreover the SCP is satisfied for a profit-maximizing firm if and only if it is also satisfied for a nonprofit firm.[42]

PROOF OF PROPOSITION 8

[41] This condition can equivalently be written as

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

[42] To see this, recall that $\Pi_p / \Pi_T = x - g_p$ where $g_p = 0$ for a nonmarketed good. For a nonprofit firm $dx^* / d\theta = -\Pi_p / \Pi_x$ and $dx^* / dT = -\Pi_p / \Pi_x$, thus $(dx^* / dp)/(dx^* / dT) = x - g_p$ and the SCP is satisfied for either type of firm if and only if it is satisfied for the other.

The unit payment that induces the first-best level of output is derived by setting $dx^* /d\theta = dx^{fb} /d\theta$ and solving for p. By the integral form of the envelope theorem, if for all $\theta \in \Theta$ we have $x^*(p, \theta) = x^{fb}(\underline{\theta}) + \int_0^{\theta} (\partial x^{fb}/\partial\tilde{\theta}) d\tilde{\theta}$, then $x^* = x^{fb}$ at every $\theta \in \Theta$.

PROOF OF PROPOSITION 9

To show that the price-inducing the first-best level of output is not second-best optimal for a marketed good (M), and is socially-optimal for a nonmarketed good (NM) we will use optimal control. The proof proceeds taking the first-order condition of Hamiltonian and showing that the FOC is not equal to zero at the price schedule identified by Proposition 8 and $x^{fb}(\theta)$ for M but is equivalent to 0 for NM. To facilitate comparison between uncertainty in the cost state and demand state we will include the state parameter in both the quality-adjusted cost and value functions. When uncertainty is in the cost state $V_\theta = 0$ and $x_\theta = 0$.

Let U be the state variable where

$$U \equiv \begin{cases} V(x; p, \theta) - g(x; p, \theta) & \text{(M)}, \\ V(x; \theta) - (1 + \gamma)g(x; \theta) & \text{(NM)}, \end{cases} \tag{A-16}$$

and let x be the control variable. The Hamiltonian for the control problem is simply $H = Uf(\theta) + \delta(\theta)d\{V - g\} /d\theta$, where δ is the Pontryagin multiplier. Therefore, for M and NM, the Hamiltonians are

$$H = \begin{cases} Uf(\theta) + \delta(\theta)\big[\big(V_x - g_x\big)\frac{dx^{fb}}{d\theta} + \big(V_p - g_p + (V_x - g_x)\frac{dx^{fb}}{dp}\big)\frac{dp^{np}}{d\theta} + V_\theta - g_\theta)\big] & \text{(M)}, \\ Uf(\theta) + \delta(\theta)\big[\big(V_x - (1 + \gamma)g_x\big)\frac{dx^{fb}}{d\theta} + V_\theta - (1 + \gamma)g_\theta)\big] & \text{(NM)}. \end{cases}$$

By the maximum principle

$$\dot{\delta} = -\frac{\partial H}{\partial U} = -f(\theta) \tag{A-17}$$

The boundary $\theta = \underline{\theta}$ is unconstrained; therefore, the transversality condition at $\theta = \underline{\theta}$ is

$$\delta(\underline{\theta}) = 0. \tag{A-18}$$

Integrating (A-17) gives

$$\delta(\theta) = -F(\theta). \tag{A-19}$$

The p that maximizes H is the p solving

(A-20)
$$dU /dp\ f(\theta) - F(\theta)\Big[\big(V_{xp} - g_{xp}\big)\tfrac{dx^{fb}}{d\theta} + V_{\theta p} - g_{\theta p} + \Big((V_{xx} - g_{xx})\big(\tfrac{dx^{fb}}{dp}\big)^2 + 2(V_{xp} - g_{xp})\tfrac{dx^{fb}}{dp} + V_{pp} - g_{pp} + (V_x - g_x)\tfrac{d^2x^{fb}}{dp^2}\Big)\tfrac{dp^{np}}{d\theta}\Big] = 0,$$

for M and

$$dU /dp\ f(\theta) - F(\theta)\big[\big(V_{xx} - (1 + \lambda)g_{xx}\big)\tfrac{dx^{fb}}{d\theta} + (V_{\theta x} - (1 + \lambda)g_{\theta x})\big] = 0, \tag{A-21}$$

for NM.

The first-best outcome requires $dU/dp = 0$ for both M and NM and the expressions in brackets are equal to $d^2\{V-g\}/dpd\theta$. Because $d\{V-g\}/dp \equiv 0$ for all $\theta \in \Theta$ we have $d^2\{V-g\}/dpd\theta \equiv 0$ at the first best price and quantity.

Starting with NM, the underlying level of quality is uniquely determined by the equilibrium output quantity, x^{fb}. Moreover, neither $V(x^*;\theta)$ nor $g(x^*;\theta)$ are direct functions of price. Therefore, if $p^{np}(\theta)$ induces $x^* = x^{fb}$ at every $\theta \in \Theta$, it must be the case that the term in brackets satisfies

$$\frac{d^2\{V-g\}}{d\theta dp}(x^{fb};\theta) = \frac{d^2\{V-g\}}{d\theta dx}(x^{fb};\theta)\frac{dx^*}{dp}(p^{np},\theta) \equiv 0.$$

Therefore $dU(\theta)/dp = 0$ and the second-best is equivalent to the first-best, which has already been shown to be equivalent to the socially optimal outcome for a nonmarketed good.

Returning to M, 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. Inducing the consumers to demand the first-best quantity requires that p^{np} solve Eq. (A-20); however, inducing the firm to supply the appropriate level of quality while maintaining incentive compatibility requires that p^{np} also satisfy

$$\text{(A-22)} \qquad p^{np}(\theta) = g_x\big(x^{fb};p^{np}(\theta),\theta\big) + \frac{g_\theta\big(x^{fb};p^{np}(\theta),\theta\big)}{dx^{fb}/d\theta} \text{ for all } \theta \in \Theta.$$

Eq. (A-20) and (A-22) are independent, thus $p^{np}(\theta) = p^{fb}(\theta)$ for all $\theta \in \Theta$ requires $dp^{np}/d\theta = dp^{fb}/d\theta$ at all $\theta \in \Theta$. However, generically we have

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

Hence, $p^{np} \neq p^{fb}$ and Eq. (A-20) is not equal to zero at x^{fb} and p^{np} given its strict concavity. Therefore $\{x^{fb}(\theta), p^{np}(\theta)\}$ is not first-best optimal for a marketed good.

References

Aguirre, Inaki and Arantza Beitia, "Regulating a Monopolist with Unknown Demand: Costly Public Funds and the Value of Private Information," *Journal of Public Economic Theory*, 2004, *6* (5), 693–706.

Allen, Robin and Paul Gertler, "Regulation and the Provision of Quality to Heterogeneous Consumers: The Case of Prospective Pricing of Medical Services," *Journal of Regulatory Economics*, 1991, *3*, 361–375.

Armstrong, Mark and David E.M. Sappington, "Toward a Synthesis of Models of Regulatory Policy Design with Limited Information," 2004. Working paper.

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

Baron, David and David Besanko, "Regulation, Asymmetric Information, and Auditing," *Rand Journal of Economics*, 1984, *15*, 447–470.

_ **and Roger Myerson**, "Regulating a Monopolist with Unknown Costs," *Econometrica*, 1982, *50*, 911–930.

Baron, David P., "Price Regulation, Product Quality, and Asymmetric Information," *American Economic Review*, 1981, *71* (1), 212–220.

Beitia, Arantza, "Hospital Quality Choice and Market Structure in a Regulated Duopoly," *Journal of Health Economics*, 2003, *22*, 1011–1036.

Caillaud, B., Roger Guesnerie, and Jean-Jaques Tirole, "Government Intervention in Production Incentives Theory: A Review of Recent Contributions," *RAND Journal of Economics Journal of Economics*, 1988, *19* (1), 1–26.

Chalkley, Martin and James M. Malcomson, "Contracting for health services when patient demand does not reflect quality," *Journal of Health Economics*, 1998, *17*, 1–19.

_ **and** _ , "Contracting for Health Services with Unmonitored Quality," *Economic Journal*, 1998, *108*, 1093–1110.

Colombo, John D., "The Failure of Community Benefit," *Health Matrix: Journal of Law-Medicine*, 2005, *15:29*, 29–65.

Dasgupta, Partha, P. Hammond, and Eric Maskin, "The Implementation of Social Choice Rules: Some General Results on Incentive Compatibility," *Review of Economic Studies*, 1979, *46*, 185–216.

Davis, K., "Economic Theories of Behavior in Nonprofit, Private Hospitals," *Economic and Business Bulletin*, 1972, *24* (2), 1–13.

Ellis, Randall P. and Thomas G. McGuire, "Provider Behavior Under Prospective Reimbursement," *Journal of Health Economics*, 1986, *5*, 129–151.

Frank, R. G. and D. S. Salkever, "The Supply of Charity Services by Nonprofit Hospitals: Motives and Market Structure," *RAND Journal of Economics*, 1991, *22* (3), 430–445.

Gowrisankaran, Gautam and Robert J. Town, "Competition, Payers, and Hospital Quality," *Health Services Research*, 2003, *38* (6 pt 1), 14031422.

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

Guesnerie, Roger and Jean-Jaques Laffont, "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*, 1984, *25*, 329–369.

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

Laffont, Jean-Jaques and Jean Tirole, "Using Cost Observation to Regulate Firms," *Journal of Political Economy*, 1986, *94*, 614–641.

_ **and** _ , *A Theory of Incentives in Procurement and Regulation*, MIT Press, 1993.

Landon, Bruce E., Charon-Lise T. Normand, Adam Lessler, James O'Malley, Stephen Schmaltz, Jerod L. Loeb, and Barbara J. McNeil, "Quality of Care for the Treatment of Acute Medical Conditions in U.S. Hospitals," *Archives of Internal Medicine*, 2006, *166*, 2511–2517.

Lewis, Tracy R. and David E. Sappington, "Regulating a Monopolist with Unknown Demand," *American Economic Review*, 1988, *78* (5), 986–998.

— **and David E.M. Sappington**, "Regulating a Monopolist with Unknown Demand and Cost Functions," *RAND Journal of Economics*, 1988, *19* (3), 438–457.

— **and Devid E.M. Sappington**, "Incentives for Conservation and Quality-Improvement by Public Utilities," *American Economic Review*, 1992, *82* (5), 1321–1340.

Ma, Ching-To Albert, "Health Care Payment Systems: Cost and Quality Incentives," *Journal of Economics and Management Strategy*, 1994, *3* (1), 93–112.

Myerson, Roger, "Incentive Compatibility and the Bargaining Problem," *Econometrica*, 1979, *47*, 61–73.

National Center for Charitable Statistics, "The Nonprofit Sector in Brief," Technical Report, Urban Institute 2007.

Needleman, Jack, "The Role of Nonprofits in Healthcare," *Journal of Health Politics, Policy and Law*, 2001, *26*, 1113–1130.

Newhouse, Joseph P., "Toward a Theory of Nonprofit Institutions: An Economic Model of a Hospital," *American Economic Review*, 1970, *60* (1), 64–74.

Oleck, Howard L. and Martha E. Stewart, *Nonprofit Corporations, Organizations, and Associations*, 6th ed., Prentice Hall Trade, 1994.

Owens, Bramer, "The Plight of the Not-for-Profit," *Journal of Healthcare Management*, 2005, *50* (4), 237–250.

Pauly, Mark and M. Redisch, "The Non-Profit Hospital as Physician Cooperative," *American Economic Review*, 1973, *63*, 87–100.

Riordan, Michael, "On Delegating Price Authority to a Regulated Firm," *Rand Journal of Economics*, 1984, *15*, 108–115.

Rogerson, William P., "On the Optimality of Menus of Linear Contracts," 1987. Northwestern Discussion Paper #714R.

— , "Choice of treatment intensities by a nonprofit hospital under prospective pricing," *Journal of Economics and Management Strategy*, 1994, *3*, 7–51.

Rose-Ackerman, S., "Ideal Versus Dollars: Donors, Charity Managers, and Government Grants," *Journal of Political Economy*, August 1987, *95*, 810–823.

Sappington, David E.M., "Optimal Regulation of of Research and Development under Imperfect Information," *Bell Journal of Economics*, 1982, *13*, 354–368.

Spence, A. Michael, "Monopoly, Quality, and Regulation," *Bell Journal of Economics*, 1975, *6* (2), 417–429.

Wolinsky, Asher, "Regulation of Duopoly: Managed Competition vs Regulated Monopolies," *Journal of Economics and Management Strategy*, 1997, *6* (4), 821–847.