CONTRACTING FOR PUBLICLY FUNDED SERVICES: UNMONITORED QUALITY AND EFFICIENCY*

KEVIN E. PFLUM†

ABSTRACT. I analyze the optimal regulatory policy for a setting in which the service is provided by regulator funds and the contracted firm has superior knowledge of demand, which it can manipulate through its choice of quality. I show the conditions with which the firm can be induced to use its private information in the social interest. Because the firm can manipulate demand, however, the firm will generally extract an information rent, which is independent of the social cost of public funds. I further show how the firm's objectives impact the optimal contract and what the regulator can achieve. The findings provide new insights into the conditions that lead to market distortions when there is asymmetric information.

JEL Codes: H42, H57, L20, L31

Key Words: asymmetric demand information, regulating quality, for-profit and not-for-profit

1 Introduction

It is well known that a regulated firm will extract an information rent resulting in a downward distortion in output when it has superior knowledge of its costs (see, for example, Myerson, 1979; Baron and Besanko, 1984; Laffont and Tirole, 1986). However, this result does not carry over to all situations in which the firm may have an information advantage. For example, Lewis and Sappington (1988) show that there will be no distortion from first best when a regulated firm has increasing marginal costs and superior knowledge of demand; and, Caillaud, Guesnerie, and Tirole (1988) show with an example of hidden effort and unobserved firm cost that the firm will provide the efficient level of effort if it is able to internalize all of the gains from exerting effort. The firm is not able to internalize the gains, however, when it is reimbursed based on observable costs and effort will be 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.[1]

The purpose of this study is to add to our understanding of when asymmetric information generates market distortions. I build off of Lewis and Sappington (1988) and Aguirre and Beitia (2004)

*I am grateful to Paul J. Healy, Dan Levin, Lixin Ye, Huanxing Yang, Paula Cordero Salas, and seminar participants at the Ohio State University Microeconomic Theory Seminar for helpful comments and discussions.

†University of Alabama, Department of Economics, Finance and Legal Studies, Box 870224, Tuscaloosa, AL, 35406, kpflum@cba.ua.edu.

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

by examining a regulator's problem when the contracted firm has superior knowledge of its demand[2] that it can additionally manipulate through its choice of quality.[3] Because many services for which quality is typically an important factor (e.g., health care and education) are frequently funded either completely, or almost completely, through public funds I focus on the latter case in which payment is made only using public funds.[4] This restriction is important because when consumers are responsible for the payments, changes in the unit payment will impact the quantity demanded both directly and indirectly as the firm adjusts its choice of quality based on the payment it receives. In consequence, a regulator will not be able to induce the social optimum—even if it knows the demand state (Spence, 1975; Baron, 1981). In contrast, by using public funds the regulator has the ability to achieve the social optimum when it either has knowledge of the demand state or can observe the firm's choice of quality and we can identify how the combination of moral hazard and adverse selection in this particular environment may limit the regulator.

Understanding what the regulator can achieve when quality is added to the model is important for a couple reasons. First, policy makers have long been concerned with the quality of care provided in health care markets and how it can be improved. To this end different pay-for-performance programs have been developed to improve quality and the recently passed Affordable Care Act (ACA) encourages experimentation by the Centers for Medicare and Medicaid Services (CMS) to research, develop, and test new payment programs with the objective of increasing the quality of care. There have been similar concerns in education. For example, the National No Child Left Behind Act (NCLB) enacted in 2002 spends around $2.4 billion annually to generate a verifiable quality metric on which the government makes school funding and employee retention decisions. However, this expense is efficient only if the social benefit of generating the verifiable measure of quality exceeds this cost.

Second, it's unclear how much the addition of quality will impact what the regulator can optimally achieve. As access to the service is paid for via public funds, the firm's choice of quality impacts its revenue through the effect on demand only. The firm has no ability to alter the amount it charges for its service after adjusting quality and there will be no direct impact to demand when

[2] As Lewis and Sappington (1988) argue, the firm's superior knowledge of the demand state can develop because a locally present firm will have direct contact with and be composed of employees from the given market allowing it to develop superior market knowledge to that of a regulator with no local presence.

[3] Lewis and Sappington show that when the regulated firm has increasing marginal costs and consumers pay for the service the firm will not be able to extract an information rent and there is no distortion from first-best. Aguirre and Beitia (2004) extend Lewis and Sappington (1988) by having the regulator pay the fixed transfer while consumers make the unit payment. Public funds are more costly so the regulator allows the firm to extract some rents to limit the fixed transfer amount in favor of the less costly unit payments.

[4] Other studies that consider optimal payment mechanisms in which the regulated service is also paid via public funds include Ellis and McGuire (1986), Ma (1994), Chalkley and Malcomson (1998b) and Jack (2005), which all focus on regulating health care providers where demand is known but there are other sources of adverse selection such as the consumer's type or the provider's costs or degree of altruism.

the regulator adjusts the unit payment. Any distortion away from first-best will critically depend on how much the benefits of the quality choice are internalized by the firm.

In Lewis and Sappington (1988) the regulator can achieve first-best because it knows with certainty what the firm's marginal cost of production will be for some announcement of the demand state. That the firm can manipulate demand with its choice of quality would not pose a problem for the regulator if it knew the demand state because it would know what is efficient. However, although it has knowledge of firm's cost of production, since it does not know the true demand state, it is uncertain about what the firm's quality-adjusted marginal costs should be. In other words, the regulator's problem is *as if* it did not know the firm's technology and the firm will extract an information rent similar to that found in studies focussing on a setting where the firm's private information is with respect to its costs (e.g., Baron and Myerson, 1982; Baron and Besanko, 1984; Laffont and Tirole, 1986) and both quality and output will be distorted downwards.[5] There will not be any distortion, however, when the quality-adjusted marginal cost of production does not change with the demand state. In this case the firm still extracts an information rent as its total cost of production differs with the demand state so the contract must satisfy incentive compatibility, but the rent it can extract is independent of the quantity it produces in this case so the regulator can optimally induce the efficient level of quality and output.

I also analyze how the firm's objectives impact what the regulator can achieve by analyzing the problem for a firm that seeks to maximize consumer benefit subject to a zero profit constraint.[6] This is important because many of the relevant markets such as health care and education also exhibit a mix of for-profit and not-for-profit firms.[7] Although a firm that maximizes consumer surplus has no preference for profit, its objectives are not perfectly aligned with the regulator's and any payment policy that the regulator offers is still constrained by incentive compatibility. Nevertheless, the regulator can (mostly) ignore the firm's profit giving it an extra degree of freedom to focus on ensuring the firm's choice of quality is socially efficient. I provide the conditions with which a regulator can induce the nonprofit firm to use its private information to produce in the full social interest.

The remainder of the paper is organized as follows. In Section 2 the basic model is developed. Section 3 examines the regulator's problem when the firm is profit-maximizing and the section is

[5]This differs from the much more complicated setup in Lewis and Sappington (1992) in which the regulator has uncertainty regarding both the demand state and the firm's cost of production as the current model is still a one-dimensional screening problem.

[6]Ellis and McGuire (1986) consider the regulatory policy for a firm that is not profit maximizing. They show that when the preferences of the firm are the same as the regulator's, the appropriate payment rule will induce the firm to produce the socially preferred level of quality but, if the firm has less of a preference for consumer benefits, then the regulator must subsidize the firm in order to induce the socially preferred level of quality. Chalkley and Malcomson (1998a) and Jack (2005) also allow the regulated firm to have altruistic motives. The main difference between these models and the current paper is that demand is inelastic to quality—requiring altruism if the firm is to provide any quality.

[7]For example, of the 4,897 registered community hospitals in the United States, 873 are for-profit while 2,913 are non-profit, and twelve states permit for-profit corporations to operate charter schools.

divided into three parts. The first part establishes the baseline contract when information is symmetric and shows that there is a unique contract inducing the first-best outcome. In the second part, I show that the result in Lewis and Sappington (1988) breaks down when the firm can manipulate demand, even if the regulator can contract directly on verifiable output. The third part derives the optimal payment policy. Section 4 examines the regulator's problem when the firm is not-for-profit and seeks to maximize gross consumer surplus. Section 5 ends with some final remarks and conclusions. Unless otherwise presented, proofs can be found in the appendix.

2 The Model

Consider a market environment in which there is a single firm supplying a service that the regulator would like to provide to society at no direct cost.[8] The firm provides the service at some level of quality $q \geq 0$. Quality is observable by consumers but not verifiable so cannot be directly contracted upon. Consumer demand, $x(q, \theta)$, is a function of the level of quality q and the demand state $\theta \in \Theta = [\underline{\theta}, \overline{\theta}]$, and it is a measure of the firm's output such as the number of students enrolled or patients treated. Consumers need not perfectly observe quality as long as demand responds to changes in quality; i.e., $x_q(\cdot) \neq 0$. Demand is a twice continuously differentiable function, increasing in both q and θ, and strictly concave in q. Furthermore it is assumed that $x(0, \theta) = 0$, for all $\theta \in \Theta$ so that the firm must supply some positive level of quality in all demand states. When the firm has superior knowledge of the state θ, 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.The characteristics of the regulator's uncertainty are common knowledge.

Let $q(x, \theta)$ denote the inverse demand function, that is, $q(x, \theta)$ denotes the level of quality required to induce the demand x given the demand state is θ. Observe that the properties of $x(q, \theta)$ are sufficient to insure the existence of $q(x, \theta)$. An intuitive way of viewing the firm's problem is that it selects the level of q maximizing its objective; however, it is equivalent to view the firm as choosing the quantity which maximizes its objective given that it must set the quality, $q(x, \theta)$, in order to induce a demand for the chosen quantity x.[9]

---

[8]The market need not be a monopoly. The demand function can be interpreted as the firm's residual demand given the other firms, which are regulated, provide the equilibrium level of quality. Many papers have examined the use of competition to increase quality in similar market environments; e.g., Ma and Burgess (1970), McGuire and Riordan (1995), Wolinsky (1997) and more recently Beitia (2003). See also Gaynor (2006) for a survey of the issues concerning the nature of competition in health care markets and the impact of competition on quality.

[9]The firm is assumed to not have a binding capacity constraint. If it did, then the relationship between quantity and quality breaks down once the capacity limit is reached. See Chalkley and Malcomson (1998b) for an example of a regulatory policy inducing the desired level of quality when the firm has capacity constraints. Note, however, that the result in Chalkley and Malcomson (1998b) critically depends on the regulator knowing not just the quantity of service provided by the firm, but the demand the firm faces.

The cost of producing quantity x at quality q is given by the function $c(x, q)$. The cost of production is assumed to be thrice continuously differentiable, strictly increasing over $[0, \infty)^2$, and strictly convex in x. Similar to Rogerson (1994),[10] define $g(x, \theta)$ as the firm's *quality-adjusted* cost function

$$g(x, \theta) = c\big(x, q(x, \theta)\big). \tag{1}$$

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

$$\frac{dg}{dx}(x, \theta) = \frac{dc}{dx}\big(x, q(x, \theta)\big) = \frac{\partial c}{\partial x}\big(x, q(x, \theta)\big) + \frac{\partial c}{\partial q}\big(x, q(x, \theta)\big)\frac{dq}{dx}(x, \theta). \tag{2}$$

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 more demand given the demand state.

Given the properties of $c(x, q)$ and $q(x, \theta)$, $g(x, \theta)$ must be thrice continuously differentiable. I further assume that $g(\cdot)$ is strictly increasing and convex in x and that the quality-adjusted marginal costs are non-increasing with the demand state, $g_{x\theta}(\cdot) \leq 0$, weakly convex in the demand state, $g_{x\theta\theta}(\cdot) \geq 0$, and the change in costs with the demand state are weakly concave in quantity, $g_{\theta xx}(\cdot) \leq 0$. The property $g_{x\theta}(\cdot) \leq 0$ has two implications. First, it implies that marginal costs are weakly increasing with quality because, *ceteris paribus*, an increase in the demand state results in a reduction in the service quality, $q_\theta(\cdot) < 0$. Second, $g_{x\theta}(\cdot) \leq 0$ implies that the firm prefers to supply a weakly higher quantity of service in higher demand states. The second and third properties, $g_{x\theta\theta}(\cdot) \geq 0$ and $g_{\theta xx}(\cdot) \leq 0$, are for technical reasons and simplify the analysis of the second-best contract; however, they are not necessary for the results.

The regulator values the benefits that the consumers receive according to the function $B(x, q, \theta)$, which may reflect the consumers' direct value of consumption (i.e., $B(x, q) = \int_0^x P(\tilde{x}, q)d\tilde{x}$ where P is inverse demand) or the regulator's valuation of consumption in the presence of social externalities as are common with services such as health care and education. The benefit to consumers is increasing in the quantity consumed, the quality of the service, and the demand state ($B_x(\cdot) > 0$, $B_q(\cdot) > 0$, and $B_\theta(\cdot) > 0$). B is strictly concave in q and thrice continuously differentiable. Again, it will be convenient to define $V(x, \theta)$ as the *quality-adjusted* benefit function. As with g, $V(x, \theta)$ denotes the benefits the consumers receive given that the quality has been adjusted to induce a demand of x when the demand state is θ. The properties of B imply V is strictly concave in x and thrice continuously differentiable. It is also assumed that $V_{x\theta}(\cdot) \geq 0$ and $\lim_{x \to 0} V(x, \theta) = -\infty$ for all $\theta \in \Theta$. The first property is a regularity condition insuring that the equilibrium level of quality does not drop so much with an increase in demand that the net effect is to lower social

[10] In Rogerson (1994) demand is deterministic so there is no adverse selection parameter θ in g.

surplus in a higher demand state, and the second property ensures that it is always socially optimal to provide some positive quantity of service.

The regulator is a Stackelberg leader and is endowed with the power to establish a unit price $p \in \mathbb{R}_+$ and transfer payment $T \in \mathbb{R}$.[11] Given a payment policy $\{p, T\}$, the firm's profit is expressed as

$$\Pi(p, T; \theta) = px - g(x, \theta) + T. \tag{3}$$

The for-profit firm has the objective of choosing the quantity which maximizes profits and the not-for-profit firm is assumed to be completely altruistic with the objective of maximizing consumer benefits subject to a non-negative profit constraint.[12]

The regulator is responsible for paying both p and T and there is an opportunity cost of raising public funds denoted as $\lambda > 0$. Thus every \$1 raised by the regulator has a social cost of $\$(1 + \lambda)$. Total consumer surplus is the social benefit of the service provided by the firm minus the social cost of the service

$$CS(p, T; \theta) = V(x, \theta) - (1 + \lambda)(px + T). \tag{4}$$

Regardless of the profit-status of the firm, the regulator's objective is to maximize total social surplus:

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

Denote $\hat{x}$ as the quantity maximizing the firm's objective. By substituting Π and CS into (5) and rearranging, the regulator's problem can be expressed as

$$\max_{p,\Pi} \; V(\hat{x}, \theta) - (1 + \lambda)g(\hat{x}, \theta) - \lambda\Pi \;\text{ such that } \Pi \geq 0. \tag{6}$$

Because there is a shadow cost to public funds, the regulator's objective function reaches its maximum only if the firm earns zero profit. Furthermore, the first order condition with respect to p identifies the efficient quantity x^* as the $x > 0$ solving

$$V_x(x^*, \theta) = (1 + \lambda)g_x(x^*, \theta). \tag{7}$$

Putting the preceding arguments together the first-best outcome can be formally defined.

Definition 1. *The first-best outcome is the outcome in which the firm earns zero profit and the equilibrium quantity is the socially efficient quantity:*

$$x^* = \arg\max_x V(x, \theta) - (1 - \lambda)g(x, \theta).$$

[11] Both proposition 2 and lemma 2 show that the results are robust to other contractual forms.

[12] The nonprofit firm's objective could instead be to maximize some convex combination of profit and social value similar to Ellis and McGuire (1986); Chalkley and Malcomson (1998a); Jack (2005). The firm's behavior is only different when its profit constraint binds, however, since otherwise it will still behave as a marginal optimizer producing the level of output that equates the quality-adjusted marginal cost weighted by its preference for profit with the unit price (Pflum, 2011).

3 Regulating a Profit-Maximizing Firm

3.1 The Full-Information Policy

It is assumed that the regulator and firm have symmetric information regarding all aspects of the model. A contract $\{p, T\}$ is said to induce x^* if x^* is the solution to the firm's optimization program,

$$x^* = \hat{x} = \arg\max_x px - g(x,\theta) + T. \tag{8}$$

The first order condition from the firm's problem shows that the firm's optimizer $\hat{x}(p,\theta)$ is the unique

$$p = g_x(\hat{x},\theta) = dc\big(\hat{x}, q(\hat{x},\theta)\big)\,/dx\ . \tag{9}$$

Using the definition of the first-best outcome (7) and the firm's choice of output given p, the optimal payment policy for a profit-maximizing firm can be easily found. The following lemma formally states the optimal payment policy.

Lemma 1. *The optimal contract for a profit-maximizing (FP) firm under symmetric information consists of the unique unit price $p^*_{FP}(\theta)$ and transfer payment $T^*_{FP}(\theta)$ satisfying:*

$$p^*_{FP}(\theta) = g_x\big(\hat{x}\big(p^*_{FP}(\theta),\theta\big),\theta\big) = g_x(x^*,\theta) = \tfrac{1}{1+\lambda}V_x(x^*,\theta), \tag{10}$$

$$T^*_{FP}(\theta) = g(x^*,\theta) - p^*_{FP}(\theta)x^*, \tag{11}$$

for all $\theta \in \Theta$.

The regulator induces the first-best outcome by taking advantage of the firm's optimal choice of x identified in (9). Observe that both the unit and transfer payments are needed to meet the two conditions of efficiency. The unit payment is used to induce the efficient quantity x^* while the transfer payment is required to hold the firm to zero profit. The transfer payment may be positive or negative depending on whether or not the quality-adjusted marginal cost at the efficient quantity exceeds the average cost.

3.2 Contracting with Asymmetric Information

Lewis and Sappington (1988) find that when the firm has increasing marginal costs the firm cannot take advantage of its superior knowledge of the demand state to extract rents and the first-best outcome is achieved. That is, adverse selection is not a sufficient condition for a market distortion in the second-best. In the current model the cost of production is similarly assumed to be convex. Because the firm can alter demand through its choice of quality, however, convexity in the quality-adjusted cost function is no longer sufficient and the first-best outcome can only be induced if the firm's cost of producing a given output does not optimally change across demand states. The following proposition reports this finding.

Proposition 1. *Suppose the firm's cost are strictly increasing in x and $g_{xx}(\cdot) \geq 0$, then the first-best policy reported in lemma 1 implements the first-best when there is asymmetric demand information if and only if $g(x, \theta_1) = g(x, \theta_2)$ for all $x \geq 0$ and for all $\theta_1, \theta_2 \in \Theta$.*

The reason $g_\theta(\cdot) = 0$ is needed to induce the first-best outcome follows from how the unit price is used to induce the first-best quality. When $g_\theta(\cdot) = 0$ there is only one quality-adjusted cost curve for all demand states and the regulator can set the unit price at the appropriate marginal cost. When $g_\theta(\cdot) \neq 0$, however, the quality-adjusted cost curve is unique to the demand state and the same quantity in different states will possess different quality-adjusted marginal costs.

Recall that $g_\theta(\cdot)$ is equivalent to the following

$$\frac{\partial g}{\partial \theta}\big(\hat{x}(p(\theta), \theta), \theta\big) = \frac{\partial c}{\partial q}\big(\hat{x}(p(\theta), \theta), q(\hat{x}, \theta)\big)\frac{\partial q}{\partial \theta}(\hat{x}, \theta). \tag{12}$$

From (12) it is clear that if a change in quality does not impact its cost of production or the firm does not optimally alter its quality (because demand is inelastic to quality, e.g.), then $g_\theta(\cdot)$ is zero and the regulator can induce the first-best outcome. Formally, by applying the relationship in (12) to proposition 1 we have the following corollary.

Corollary 1. *Suppose the firm's cost are strictly increasing in x and $g_{xx}(\cdot) \geq 0$, then the first-best policy in lemma 1 is a feasible solution to the regulator's problem if and only if either $\frac{\partial c}{\partial q} = 0$ or $\frac{\partial q}{\partial \theta} = 0$.*

It's notable that Lewis and Sappington (1988) assume that output cannot be observed, so cannot be contracted upon. They argue that output may be unobservable because there are different quality-adjusted levels of output and the regulator cannot observe the true quality-adjusted level. However, per corollary 1, their result only holds then if the quality-adjusted marginal costs of production are the same for each demand state. The assumption that the regulator cannot contract on output is made because otherwise it could trivially induce the first-best by offering a contract $\{P(\theta), x(\theta)\}_{\theta \in \Theta}$ where $P(\theta)$ is a fixed budget and $x(\theta)$ is output, both in state θ. Once quality is formally added to the model, however, the ability to contract directly on output is irrelevant. The reason is because the regulator still will not know what the efficient level of output should be since it does not know the demand state. In consequence, the firm will continue to be able to extract an information rent whenever $g_\theta(\cdot) \neq 0$. The following proposition formally states this observation.

Proposition 2. *Suppose the firm's cost are strictly increasing in x and $g_{xx}(\cdot) \geq 0$, then the regulator can induce it to produce the first-best level of quality through a fixed-budget and quantity contract, $\{P(\theta), x(\theta)\}_{\theta \in \Theta}$, where x is the contracted quantity and P is the firm's fixed-budget, if and only if $g_\theta(\cdot) = 0$.*

3.3 The Optimal Incentive Compatible Contract

The regulator's objective is to design a menu of payments $\{p(\theta), T(\theta)\}_{\theta\in\Theta}$ that induce the firm to truthfully reveal the demand state and maximize total expected social surplus. To achieve this the menu of prices must satisfy individual rationality so that the firm is a willing participant and incentive compatibility so that the firm is induced to truthfully announce its private information (Dasgupta et al., 1979; Myerson, 1979). The derivation of the optimal contract is standard in the contracting literature (e.g., see Guesnerie and Laffont, 1984; Caillaud, Guesnerie, and Tirole, 1988), however, the steps are provided for clarity.

Because the for-profit firm's objective is to select the contract that maximizes its profit, its *value function* is defined as

$$U(\hat{\theta}, \theta) = p(\hat{\theta})\hat{x}\big(p(\hat{\theta}), \theta\big) - g(\hat{x}, \theta) + T(\hat{\theta}), \tag{13}$$

where $\hat{\theta}$ is the firm's announcement, θ is the true demand state, and $\hat{x}$ is the firm's maximizer. With the firm's value function, the regulator's problem may be expressed as[13]

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

subject to

$$U(\theta) \geq 0 \; \forall\theta \in \Theta \qquad \text{(Individual Rationality)}$$
$$U(\theta,\theta) \geq U(\hat{\theta},\theta) \; \forall\hat{\theta},\theta \in \Theta, \qquad \text{(Incentive Compatibility),}$$

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

The following lemma characterizes the necessary and sufficient conditions for an incentive compatible payment policy $\{p(\theta), T(\theta)\}_{\theta\in\Theta}$.

Lemma 2. *The payment policy $\{p(\theta), T(\theta)\}_{\theta\in\Theta}$ is incentive compatible if and only if it satisfies the following conditions*

(i) $\dfrac{dU(\theta)}{d\theta} = \dfrac{\partial U(\theta)}{\partial\theta} = -g_\theta\big(\hat{x}\big(p(\theta),\theta\big),\theta\big) \geq 0,$

(ii) $\dfrac{dp}{d\theta} \geq 0.$

Condition *(i)* is an application of the envelope theorem and states that the change in the firm's profits across demand states must equal $-g_\theta(\cdot) > 0$. Observe that this condition is independent of the mechanism so must be satisfied for any contractual form (e.g., a fixed-budget and quality contract) and is a consequence of the fact that the firm will lower the level of service quality as the demand state increases. The resulting lower quality-adjusted marginal cost creates an incentive for

[13] Note that the properties of $g(\cdot)$ and $x(\cdot)$ are sufficient to ensure that the firm's marginal rate of substitution (MRS) of price for transfer payment (U_p/U_T) is monotonic in θ for all $\theta \in \Theta$, which is required for full type separation.

the firm to misreport so the regulator must design the payments so that the firm receives an information rent under truthful revelation. Confirming propositions 1 and 2, an incentive compatible contract must leave the firm with rents whenever $g_\theta(\cdot) \neq 0$.

Condition *(ii)* is specific to the contractual form and reports that any incentive compatible menu of prices must include a unit payment that is increasing with the demand state. This condition is a consequence of the firm's preference for increasing output in higher demand states.[14]

Lemma 2 indicates that the firm's utility is increasing in the demand state. Thus, condition *(i)* of lemma 2 implies that the firm's value function can be expressed as

$$U(\theta) = U(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} \frac{\partial U(\tilde{\theta})}{\partial \theta}\, d\tilde{\theta}. \tag{14}$$

Because the firm's rents are increasing with the demand state, individual rationality binds in only the low state $\underline{\theta}$ and the regulator can design the policy so that $U(\underline{\theta}) = 0$.

By plugging (14) into the regulator's objective function (and integrating by parts)[15], the regulator's problem can be rewritten as

$$\max_{p(\theta)} \int_{\underline{\theta}}^{\overline{\theta}} \left\{ V(\hat{x}, \theta) - (1+\lambda) g(\hat{x}, \theta) + \lambda \left(\frac{1 - F(\theta)}{f(\theta)} \right) \frac{\partial g(\hat{x}, \theta)}{\partial \theta} \right\} \, dF(\theta).$$

subject to $dp/d\theta \geq 0$. The first-order condition of the regulator's optimization program yields

$$V_x(\hat{x}, \theta) - (1+\lambda) g_x(\hat{x}, \theta) = -\lambda \left(\tfrac{1-F(\theta)}{f(\theta)} \right) g_{x\theta}(\hat{x}, \theta). \tag{15}$$

The quantity x solving eq. (15) is the second-best quantity given the regulator's constraints and the interpretation is as follows. Increasing the unit payment in demand states $[\theta, \theta + d\theta]$, which number $f(\theta)d\theta$, by dp increases the social value of consumption by $\big(V_x(\hat{x}, \theta) - (1 + \lambda) g_x(\hat{x}, \theta)\big) x_p dp$. However, from (i) of lemma 2, the increase in output simultaneously increases the firm's rent in all demand states $[\theta, \overline{\theta}]$, which number $1 - F(\theta)$, by $-g_{x\theta}(\hat{x}, \theta) x_p dp$. The total social cost of the increase in the firm's rent is $-\lambda\big(1 - F(\theta)\big) g_{x\theta}(\hat{x}, \theta) x_p dp$. Eq. (15) equates the marginal social benefit of output to the marginal social cost where social cost is higher than the firm's cost of production because the firm captures a socially costly information rent.

---

[14]This preference follows from the assumption $g_{x\theta}(\cdot) \leq 0$. If instead $g_{x\theta}(\cdot) > 0$, then an incentive compatible regulatory policy would require $dp/d\theta \leq 0$.

[15]Integrating by parts yields:

$$\begin{aligned}
\int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\theta}}^{\theta} \frac{\partial g}{\partial \theta}\, d\theta dF(\theta) &= \left[F(\theta) \int_{\underline{\theta}}^{\theta} \frac{\partial g}{\partial \theta}\, d\tilde{\theta} \right] \Bigg|_{\underline{\theta}}^{\overline{\theta}} - \int_{\underline{\theta}}^{\overline{\theta}} F(\theta) \frac{\partial g}{\partial \theta}\, d\theta \\
&= \int_{\underline{\theta}}^{\overline{\theta}} \big(1 - F(\theta)\big) \frac{\partial g}{\partial \theta}\, d\theta \\
&= \int_{\underline{\theta}}^{\overline{\theta}} \left(\frac{1 - F(\theta)}{f(\theta)} \right) \frac{\partial g}{\partial \theta}\, dF(\theta).
\end{aligned}$$

Recall that the firm will choose the level of quality equating its marginal cost of production with the unit price, so by substituting p for $g_x(\hat{x}, \theta)$ in (15), we can identify the optimal payment rule reported by proposition 3.

Proposition 3. *The optimal contract for a profit-maximizing (FP) firm under asymmetric information consists of the unique unit price $p^{**}_{FP}(\theta)$ and transfer payment $T^{**}_{FP}(\theta)$ satisfying:*

$$p^{**}_{FP}(\theta) = \left(\tfrac{1}{1+\lambda}\right)\left[V_x\left(\hat{x}\left(p^{**}_{FP}(\theta), \theta\right), \theta\right) + \lambda \tfrac{(1-F(\theta))}{f(\theta)} g_{x\theta}\left(\hat{x}\left(p^{**}_{FP}(\theta), \theta\right), \theta\right)\right],$$

$$T^{**}_{FP}(\theta) = g\left(\hat{x}\left(p^{**}_{FP}(\theta), \theta\right), \theta\right) - p^{**}_{FP}(\theta)\hat{x}\left(p^{**}_{FP}(\theta), \theta\right) - \int_{\underline{\theta}}^{\theta} g_\theta\left(\hat{x}\left(p^{**}_{FP}(\tilde{\theta}), \tilde{\theta}\right)\right) d\tilde{\theta}$$

for all $\theta \in \Theta$.

As when the regulator has complete information, with incomplete information the unit payment is used to induce the firm to produce the desired output while the transfer is used to fix the firm's profit. In contrast to when the regulator has complete information, the quantity will generally be distorted below the socially efficient level. This occurs because the firm's superior information allows it to extract an information rent making it too socially costly to produce at the first-best level. As is typical in screening models there is no distortion in output for the highest type, however, the highest type also extracts the most information rent while the distortion is highest for the lowest type, which does not earn any information rent. Several additional remarks can be made regarding the optimal payment policy.

Remark 1. The shadow cost of public funds plays a role in the equilibrium outcome, but unlike Aguirre and Beitia (2004) not in the rent the firm attains. When the shadow-cost is zero ($\lambda = 0$), the unit payment induces the firm to produce the quantity equating the social marginal benefit to the quality-adjusted marginal cost. However, the firm still attains its information rent as captured by the last term of the transfer payment, $T^{**}(\theta)$ because the contract must still be incentive compatible. Aguirre and Beitia (2004) find that the shadow cost impacts the firm's rents because the lump-sum payment is relatively more expensive so the regulator prefers to use the unit payment slightly more. When the shadow cost is zero, the unit price equals the firm's marginal cost and the firm does not earn an information rent.

Remark 2. Confirming the statement of proposition 1, the payment policy induces the first-best outcome if and only if $g_\theta(\cdot) = 0$. Recall that the first-best outcome includes two conditions: (i) the equilibrium quantity is the efficient quantity, and (ii), the firm is held to zero profit. When $g_\theta(x, \theta) = 0$, the unit payment equals the efficient unit payment—even for a non-zero shadow cost of public funds—therefore the firm will produce the efficient quantity. The firm is also held to zero profits because the last term of T^{**}_{FP} is zero when $g_\theta(\cdot) = 0$.

Remark 3. The equilibrium quantity will be efficient when the firm's quality-adjusted marginal cost does not vary with the demand state; i.e., when $g_{x\theta}(x, \theta) = 0$. In this case, the rents that the firm will extract are not a function of the quantity that it produces. In consequence, the regulator does not need to induce a lower quality level in order to limit the firm's rents. However, the firm still has an information advantage so the transfer payment must be set appropriately to satisfy incentive compatibility leaving the firm with an information rent.

4 Regulating a Nonprofit Firm

4.1 The Full-Information Policy

As with the analysis for the profit-maximizing firm, it is assumed that the regulator and firm have symmetric information regarding all aspects of the model and that the quality of service is not verifiable in a way that would allow the firm and regulator to contract directly on the quality. The nonprofit firm's objective is to maximize consumer value (by maximizing total output), subject to earning non-negative profits. The firm's optimization program is defined as

$$\max_{x>0} V(x,\theta) \text{ subject to } \Pi(x;p,T,\theta) \geq 0. \tag{16}$$

The fact that the firm limits itself to zero profit grants the regulator tremendous flexibility in how it sets the payments. The regulator simply needs to ensure that the firm's total compensation equals the cost of producing the socially efficient output and that the unit payment is no greater than the firm's quality adjust marginal cost at the socially efficient output level. The first condition is immediately obvious, while the second condition follows from the fact that profit is increasing with output as long as the unit price is higher than the marginal cost. The following proposition formally defines the set of contracts that induce the socially efficient outcome.

Proposition 4. *When the regulated firm is an output-maximizer, the first-best policy consists of any pair* $\{p(\theta), T(\theta)\} \in \Omega$ *where*

$$\Omega = \big\{\{p(\theta), T(\theta)\} \mid 0 \leq p(\theta) \leq g_x(x^*,\theta) \textit{ and } T(\theta) = g(x^*,\theta) - p(\theta)x^*\big\}.$$

4.2 Contracting with Asymmetric Information

Similar to a profit-maximizing firm, an output-maximizing firm has an incentive to misreport the demand state if doing so allows it to increase its objective—in this case output—so any payment rule must still be incentive compatible. Because the firm chooses the quantity giving it zero profit, an incentive compatible payment policy is the policy under which the firm's output is maximal when it announces the true demand state as the following definition formalizes.

Definition 2. *A payment policy* $\{p(\theta), T(\theta)\}_{\theta\in\Theta}$ *is incentive compatible for an output-maximizing firm when*

$$\hat{x}\big(p(\theta), T(\theta), \theta\big) \geq \hat{x}\big(p(\hat{\theta}), T(\hat{\theta}), \theta\big) \tag{17}$$

for all $\hat{\theta}, \theta \in \Theta$ *where* θ *is the true state and* $\hat{\theta}$ *is the firm's announcement.*

From this definition, it must be the case that for any θ_1 and θ_2 in Θ where $\theta_1 < \theta_2$, the following hold

$$\hat{x}\big(p(\theta_2), T(\theta_2), \theta_1\big) \leq \hat{x}\big(p(\theta_1), T(\theta_1), \theta_1\big), \tag{18}$$

$$\hat{x}\big(p(\theta_1), T(\theta_1), \theta_2\big) \leq \hat{x}\big(p(\theta_2), T(\theta_2), \theta_2\big). \tag{19}$$

Adding (18) and (19) gives

$$\hat{x}\big(p(\theta_2), T(\theta_2), \theta_2\big) - \hat{x}\big(p(\theta_1), T(\theta_1), \theta_2\big) \geq \hat{x}\big(p(\theta_2), T(\theta_2), \theta_1\big) - \hat{x}\big(p(\theta_1), T(\theta_1), \theta_1\big),$$

implying

$$\int_{\theta_1}^{\theta_2} \int_{\theta_1}^{\theta_2} \frac{d^2\hat{x}}{d\hat{\theta} d\theta} d\hat{\theta} d\theta \geq 0. \tag{20}$$

Because (20) is true for all $\theta_1, \theta_2 \in \Theta$ it implies $\frac{d^2\hat{x}}{d\hat{\theta}d\theta} \geq 0$ (almost everywhere), which is equivalent to

$$\frac{d^2\hat{x}}{dTd\theta}\frac{dT}{d\hat{\theta}} + \frac{d^2\hat{x}}{dpd\theta}\frac{dp}{d\hat{\theta}} \geq 0. \tag{21}$$

Eq. (21) can be simplified by observing that the first-order condition for truth-telling is

$$\left.\frac{d\hat{x}}{d\hat{\theta}}\right|_{\hat{\theta}=\theta} = \left.\frac{d\hat{x}}{dp}\frac{dp}{d\hat{\theta}}\right|_{\hat{\theta}=\theta} + \left.\frac{d\hat{x}}{dT}\frac{dT}{d\hat{\theta}}\right|_{\hat{\theta}=\theta} = 0. \tag{22}$$

Using (22) we can rewrite (21) as

$$\frac{\partial}{\partial\theta}\left(\frac{d\hat{x}\,/dp}{d\hat{x}\,/dT}\right)\left.\frac{dp}{d\hat{\theta}}\right|_{\hat{\theta}=\theta} \geq 0. \tag{23}$$

Eq. (23) is a special case of the condition derived in Theorem 1 of Guesnerie and Laffont (1984). Before providing the intuition for condition (23) it is helpful to note that the marginal rate of substitution between p and T can be reduced further. Observe that $d\hat{x}\,/dp = -\Pi_p/\Pi_x$ and $d\hat{x}\,/dT = -\Pi_T/\Pi_x$, therefore the marginal rate of substitution of p for T is simply the output, x. Thus, condition (23) reduces to

$$\frac{d\hat{x}}{d\theta}\left.\frac{dp}{d\hat{\theta}}\right|_{\hat{\theta}=\theta} \geq 0. \tag{24}$$

Because the marginal rate of substitution between the unit and transfer payments is x, condition (23) simply states that in any incentive compatible payment policy, the unit payment must move in the same direction as the output. A monotonic marginal rate of substitution implies the firm's indifference curves for different demand states cross only once. Given that $x(\theta)$ is chosen by the firm when the payment policy is $\{p(\theta), T(\theta)\}$, then, if $x(\theta + d\theta)$ is greater than $x(\theta)$, the payment policy $\{p(\theta + d\theta), T(\theta + d\theta)\}$ must be chosen so that $\Pi(\theta)$ is negative at $x(\theta + d\theta)$. Condition (24) indicates that any payment policy achieving this must have the characteristic $p(\theta + d\theta) \geq p(\theta)$. Similarly, if $x(\theta + d\theta) < x(\theta)$, then $p(\theta + d\theta) \leq p(\theta)$.

Using the implicit function theorem, $d\hat{x}\,/d\theta$ is derived as

$$\frac{d\hat{x}}{d\theta} = -\frac{\Pi_\theta}{\Pi_x} = -\frac{-g_\theta(\hat{x},\theta)}{p - g_x(x,\theta)} = \frac{-g_\theta(\hat{x},\theta)}{g_x(\hat{x},\theta) - p}. \tag{25}$$

Because $g_\theta(\cdot) < 0$ and $p \leq g_x(\cdot)$ it follows that $d\hat{x}\,/d\theta > 0$ and $dp(\theta)/d\hat{\theta} \geq 0$. The previous arguments lead to the following lemma.[16]

Lemma 3. *The payment policy $\{p(\theta), T(\theta)\}_{\theta\in\Theta}$ is incentive compatible if and only if it satisfies the following conditions*

(i) $\dfrac{d\hat{x}(\theta)}{d\theta} = \dfrac{-g_\theta(\hat{x},\theta)}{g_x(\hat{x},\theta) - p(\theta)} \geq 0$*,*

(ii) $\dfrac{dp(\theta)}{d\theta} \geq 0$*.*

Lemma 3 is the nonprofit analog to lemma 2. Condition *(i)* of lemma 3 identifies the path the equilibrium quantity follows across demand states. Note that the first condition of both lemmas addresses how the firm's objective—profit or output—changes with the demand state. The regulator can take advantage of condition *(i)* by setting a unit price (and transfer payment) that induces the firm's choice of output in any state to follow the socially optimal path.

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

$$\hat{x}\big(p(\theta), T(\theta), \theta\big) = \hat{x}\big(p(\underline{\theta}), T(\underline{\theta}), \underline{\theta}\big) + \int_{\underline{\theta}}^{\theta} \frac{d\hat{x}}{d\theta}\big(p(\tilde{\theta}), T(\tilde{\theta}), \tilde{\theta}\big) d\tilde{\theta}. \tag{26}$$

Therefore, by choosing a payment rule which induces the firm to choose $\hat{x}(\underline{\theta}) = x^*(\underline{\theta})$ and equates the path of $\hat{x}$ with the socially optimal path across demands states ($d\hat{x}\,/d\theta = dx^*\,/d\theta$ for all $\theta \in \Theta$), the regulator will induce the firm to choose the socially optimal output in every demand state. The following proposition identifies the optimal payment rule (when implementable).[17]

Proposition 5. *When implementable, the optimal contract for an output-maximizing (NP) firm under asymmetric information consists of the unique unit price $p^{**}_{NP}(\theta)$ and transfer payment $T^{**}_{NP}(\theta)$, where*

$$p^{**}_{NP}(\theta) = g_x(x^*(\theta),\theta) + \frac{g_\theta(x^*(\theta),\theta)}{dx^*(\theta)\,/d\theta},$$
$$T^{**}_{NP}(\theta) = g(x^*(\theta),\theta) - p(\theta)x^*(\theta).$$

Unlike when information is complete and symmetric, under asymmetric information the optimal payment rule consists of a single unit price and transfer payment for each demand state. Furthermore, similar to the payment policy for a profit-maximizing firm, the unit payment is used

[16] The proof for sufficiency can be found in the Appendix.

[17] Recall that from lemma 3 the two-part tariff is implementable only if $dp^{**}_{NP}\,/d\theta \geq 0$. Even with assumptions on the signs for all of the second and third derivatives for both V and g, $dp^{**}_{NP}\,/d\theta$ cannot be signed. See Remark 6 for more details.

to induce the firm to produce the efficient output level while the transfer payment is used to set the firm's profit. Of course the output-maximizing firm is limiting itself to zero profit regardless of the payment policy so the transfer payment is required to ensure that the firm encounters its zero profit constraint exactly at the efficient quantity. Several additional remarks can be made about this payment policy.

Remark 4. In contrast to the payment rule for a profit-maximizing firm, the optimal payment rule for a nonprofit firm is independent of the shadow cost of public funds. When regulating a profit-maximizing firm the regulator must balance efficiency with the firm's information rent and that balance depends on the social cost of the firm's profit. Because a nonprofit firm chooses the output leaving it with zero profit, however, there is nothing to balance and the shadow cost of public funds is not a factor.

Remark 5. Because $g_\theta(x^*(\theta), \theta) < 0$ and $dx^*(\theta)/d\theta > 0$ it follows that under the optimal payment rule the price is strictly less than the quality-adjusted marginal cost, $p^{**}_{NP} < g_x(x^*, \theta)$, and $dx/d\theta > 0$. However, inducing first-best may require a unit price so low as to be negative with a correspondingly high fixed transfer payment. For example, prices will be negative when the quality-adjusted marginal costs and the optimal change in x with respect to the demand state are very low while the quality adjusted-cost changes significantly with the demand state. The regulator will not be able to induce first-best in this circumstance if negative unit payments are not possible.

Remark 6. Rogerson (1987) shows that, in general a menu of linear prices cannot satisfy incentive compatibility in a similar problem for a profit-maximizing firm with unknown costs. Proposition 5 indicates that the restrictions on the social benefit and cost functions are even more limiting when the firm is output-maximizing. To see why this is the case let $\phi = dx^*(\theta)/d\theta$ and totally differentiate p^{**}_{NP} to get (with some abuse of notation):

$$\frac{dp^{**}_{NP}}{d\theta} = \frac{\phi^2 g_{x\theta} + (g_{\theta\theta} + g_{x\theta}x_\theta)\phi - (\phi_\theta + \phi_x x_p)g_\theta}{\phi^2 g_{xx} + g_{\theta x}x_p\phi - \phi_x x_p g_\theta}.$$

The expression for $dp^{**}_{NP}/d\theta$ contains a mix of second and third derivatives for both V and g. Unfortunately, there is no general way in which the higher order derivatives of V and g can be signed that will allow $dp^{**}_{NP}/d\theta$ to be unambiguously signed.[18] As a consequence, when $dp^{**}_{NP}/d\theta < 0$ there is no incentive compatible two-part tariff that induces first-best and the contract will have to take the form of a nonlinear payment.

[18]The assumptions that $g_{x\theta\theta} \geq 0$ and $g_{\theta xx} \leq 0$ in combination with concavity in the social benefit function and convexity in the firm's cost of production were sufficient to insure that the optimal contract for a profit-maximizing firm consisting of linear prices is incentive compatible.

5 Discussion and Final Remarks

This article has attempted to shed more light on the regulator's problem when the firm has superior information regarding demand. The first-best outcome has two characteristics. First, the equilibrium quantity must be the quantity equating marginal social benefit to marginal social cost; and second, the firm must be held to zero profit. By working with the quality-adjusted cost function it is clear that, when the firm alters demand through its choice of quality, the regulator's uncertainty in the demand state translates into uncertainty in what is the efficient marginal cost of production. In consequence, the regulator can only induce the first-best outcome if the level of quality does not impact the firm's cost of production. Otherwise incentive compatibility requires that the regulator leave the firm with an information rent. The second-best payment policy generally results in an inefficiently low equilibrium quantity; however, when the quality-adjusted marginal cost does not vary with the demand state the regulator will still be able to induce the efficient level of quality and quantity even though the firm extracts an information rent. Furthermore, the shadow cost of public funds impacts the output level but, in contrast to Aguirre and Beitia (2004), it does not affect the rents the firm could attain.

If instead the firm is output-maximizing, then the regulator has an extra degree of freedom as the firm solves half of the regulator's problem by leaving itself with zero profit. The regulator's problem reduces to inducing the firm to produce at the first-best level of quality. It was shown that the firm's choice of output can be controlled by the unit and transfer payments and the regulator can induce the firm to produce in the social interest. Since the regulator has no problem inducing the firm to produce the first-best level of output, the shadow cost of public funds does not create any distortions in output.

The implications of these findings are two-fold. First, even when the regulator has knowledge of the firm's technology, if the firm can affect demand though its actions, then the regulator must account for how the change in demand states impacts the firm's incentives to alter its output. Knowledge of the firm's technology as in Lewis and Sappington (1988) and Aguirre and Beitia (2004) is insufficient if the firm engages in a costly activity which varies with demand. Thus, not only is the nature of the information asymmetry important to the design of the contract, but how the firm's actions change with the information must also be identified. To reinforce this latter point, consider as a counter example the inclusion of costly effort instead of quality as the source of hidden action. Many regulatory models allow the firm to exercise costly effort that lowers the accounting costs of production. In the present model it is easy to show that, if the effect effort has on the accounting costs does not impact demand, then $g_\theta(\cdot) = 0$ and the firm will choose the first-best level of effort regardless of its objective. That is, all of the cost savings from effort will be internalized when the firm is compensated via a two-part tariff causing the firm to choose the efficient level of effort. Caillaud et al. (1988) first make this observation using an example of hidden effort and unobserved firm cost. Indeed, since the firm's level of effort does not impact demand and the adverse selection is with respect to demand the regulator will be able to induce the first-best outcome in the presence of *both* adverse selection and moral hazard. Thus, the presence of one or both is insufficient for the firm to earn an information rent and cause a distortion away from first-best.

The second implication of this study is that the regulator must carefully consider the objectives of the firm when designing the payment rules. As propositions 3 and 5 show, the optimal payment policy differs considerably depending on the objective of the firm. I have considered the two polar extremes in which a firm which is either purely self interested or purely altruistic for expositional clarity, but it is evident that if the firm's objective is to instead maximize a mix of profit and social value, then regulatory policy will need to factor in the firm's weighting if it seeks to induce the socially optimal outcome.

References

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

Armstrong, M. and D. E. Sappington (2004). Toward a synthesis of models of regulatory policy design with limited information. *Journal of Regulatory Economics 26*, 5–21. Working paper.

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

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

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

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

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

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

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

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

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

Gaynor, M. (2006). What do we know about competition and quality in health care markets? *Foundations and Trends in Microeconomics 2*(6), 441–508.

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

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

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

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

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

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

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

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

Ma, C.-T. A. and J. F. J. Burgess (1970). Quality competition, welfare, and regulation. *American Economic Review 60*, 64–74.

McGuire, T. and M. Riordan (1995). Incomplete information and optimal market structure. *Journal of Public Economics 56*, 125–141.

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

Pflum, K. E. (2011). *Issues in the Industrial Organization of Health Markets*. Ph. D. thesis, The Ohio State University.

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

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

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

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

A Mathematical Proofs

PROOF OF PROPOSITION 1: Without loss of generality let $\theta_1 < \theta_2$. Incentive compatibility (IC) requires that for any $\theta_1, \theta_2 \in \Theta$, the following hold

$$p(\theta_2)\hat{x}\big(p(\theta_2), \theta_1\big) - g\big(\hat{x}\big(p(\theta_2), \theta_1\big), \theta_1\big) + T(\theta_2) \leq 0, \text{ and} \tag{A-1}$$

$$p(\theta_1)\hat{x}\big(p(\theta_1), \theta_2\big) - g\big(\hat{x}\big(p(\theta_1), \theta_2\big), \theta_2\big) + T(\theta_1) \leq 0. \tag{A-2}$$

To simplify the notation, let $\hat{x}(\theta_j, \theta_i) = \hat{x}(p(\theta_j), \theta_i)$. By plugging the payment policy defined by (10) and (11) into (A-1) and (A-2), the IC constraints can be rewritten as

$$p\big(\theta_2\big) \geq \frac{g\big(\hat{x}(\theta_2, \theta_2), \theta_2\big) - g\big(\hat{x}(\theta_2, \theta_1), \theta_1\big)}{\hat{x}(\theta_2, \theta_2) - \hat{x}(\theta_2, \theta_1)}, \text{ and} \tag{A-3}$$

$$p\big(\theta_1\big) \leq \frac{g\big(\hat{x}(\theta_1, \theta_2), \theta_2\big) - g\big(\hat{x}(\theta_1, \theta_1), \theta_1\big)}{\hat{x}(\theta_1, \theta_2) - \hat{x}(\theta_1, \theta_1)}. \tag{A-4}$$

While convexity in the cost of output, $g_{xx}(\cdot) \geq 0$, implies the following conditions

$$\frac{dg}{dx}\big(\hat{x}(\theta_2, \theta_2), \theta_2\big) \geq \frac{g\big(\hat{x}(\theta_2, \theta_2), \theta_2\big) - g\big(\hat{x}(\theta_2, \theta_1), \theta_2\big)}{\hat{x}(\theta_2, \theta_2) - \hat{x}(\theta_2, \theta_1)}, \tag{A-5}$$

$$\frac{dg}{dx}\big(\hat{x}(\theta_1, \theta_1), \theta_1\big) \leq \frac{g\big(x(\theta_1, \theta_2), \theta_1\big) - g\big(\hat{x}(\theta_1, \theta_1), \theta_1\big)}{\hat{x}(\theta_1, \theta_2) - \hat{x}(\theta_1, \theta_1)}. \tag{A-6}$$

It is clear that if $g_\theta(x, \theta) = 0$, then $g\big(\hat{x}(\theta_2, \theta_1), \theta_1\big) = g\big(\hat{x}(\theta_2, \theta_2), \theta_1\big)$ and $g\big(\hat{x}(\theta_1, \theta_1), \theta_1\big) = g\big(\hat{x}(\theta_1, \theta_1), \theta_1\big)$ yielding (A-3)$\Leftrightarrow$(A-5) and (A-4)$\Leftrightarrow$(A-6). That is, because the firm's quality-adjusted cost function is convex, the optimal contract is incentive compatible under asymmetric information whenever $g_\theta(\cdot) = 0$ and $g_\theta(\cdot) = 0$ is sufficient.

To see that $g_\theta(\cdot) = 0$ is necessary, suppose $g_\theta(x, \theta) > 0$. Because the firm chooses the level of quality that equates its marginal cost with the unit price we have $p(\theta_2) = g_x\big(\hat{x}(\theta_2, \theta_2), \theta_2\big) = g_x\big(\hat{x}(\theta_2, \theta_1), \theta_1\big)$, which with (A-3) implies

$$\frac{dg}{dx}\big(\hat{x}(\theta_2, \theta_1), \theta_1\big) \geq \frac{g\big(\hat{x}(\theta_2, \theta_2), \theta_2\big) - g\big(\hat{x}(\theta_2, \theta_1), \theta_1\big)}{\hat{x}(\theta_2, \theta_2) - \hat{x}(\theta_2, \theta_1)}. \tag{A-7}$$

However, convexity implies

$$\begin{aligned}\frac{dg}{dx}\big(\hat{x}(\theta_2, \theta_1), \theta_1\big) &\leq \frac{g\big(\hat{x}(\theta_2, \theta_2), \theta_1\big) - g\big(\hat{x}(\theta_2, \theta_1), \theta_1\big)}{\hat{x}(\theta_2, \theta_2) - \hat{x}(\theta_2, \theta_1)} \\ &< \frac{g\big(\hat{x}(\theta_2, \theta_2), \theta_2\big) - g\big(\hat{x}(\theta_2, \theta_1), \theta_1\big)}{\hat{x}(\theta_2, \theta_2) - \hat{x}(\theta_2, \theta_1)},\end{aligned}$$

which contradicts (A-7). A similar contradiction occurs when $g_\theta(x, \theta) < 0$ therefore when $g_{xx}(x, \theta) \geq 0$ the solution to the regulator's problem is implementable only if $g_\theta(\cdot) = 0 \ \forall \theta \in \Theta$.

PROOF OF PROPOSITION 2: Let a contract consist of a fixed budget $P(\theta)$ if the firm provides the quantity $x(\theta)$, and zero otherwise.[19] Any regulatory policy achieving a first-best must be equivalent to a direct revelation mechanism where the firm truthfully announces the demand parameter θ. The regulator will then set the transfer to a level inducing the firm to supply the first-best level of output and which leaves the firm with zero profit, that is the regulator chooses $P(\theta) = g\big(x^{FB}(\theta), \theta\big)$, where $x^{FB}(\theta) \equiv \arg\max_x V(x, \theta) - (1+\lambda) g(x, \theta)$.

If truth-telling is the optimal announcement strategy, then the following must be true for all $\theta_1, \theta_2 \in \Theta$ where $\theta_2 > \theta_1$:

$$P(\theta_2) - g\big(x(\theta_2), \theta_1\big) \leq P(\theta_1) - g\big(x(\theta_1), \theta_1\big) = 0, \text{ and} \tag{A-8}$$

$$P(\theta_1) - g\big(x(\theta_1), \theta_2\big) \leq P(\theta_2) - g\big(x(\theta_2), \theta_2\big) = 0. \tag{A-9}$$

Suppose $g_\theta(x, \theta) > 0$, then from (A-8) and the fact that $g_x(\cdot) > 0$ it has to be the case that

$$P(\theta_1) - g\big(x(\theta_1), \theta_1\big) > P(\theta_2) - g\big(x(\theta_2), \theta_1\big),$$

which contradicts first-best since the firm must earn zero profit in every demand state. A similar argument follows if $g_\theta(x, \theta) < 0$ and, by the contrapositive, $g_\theta(x, \theta) = 0$ is sufficient.

Furthermore (A-8) and (A-9) jointly imply

$$g\big(x(\theta_1), \theta_2\big) \geq g\big(x(\theta_1), \theta_1\big), \tag{A-10}$$

$$g\big(x(\theta_2), \theta_1\big) \geq g\big(x(\theta_2), \theta_2\big). \tag{A-11}$$

Both (A-10 and (A-11) are true for all if and only if $g_\theta(x, \theta) = 0$ and $g_\theta(x, \theta) = 0$ is necessary.

PROOF OF LEMMA 2: Recall the firm's utility function is defined as

$$U(\hat{\theta}, \theta) = p(\hat{\theta}) \hat{x}\big(p(\hat{\theta}), \theta\big) - g\big(\hat{x}\big(p(\hat{\theta}), \theta\big), \theta\big) + T(\hat{\theta}). \tag{A-12}$$

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

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

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

$$\left.\frac{\partial U}{\partial \hat{\theta}}\right|_{\hat{\theta}=\theta} = \frac{dp}{d\theta} \hat{x}\big(p(\theta), \theta\big) + \frac{dT}{d\theta} = 0. \tag{A-13}$$

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

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

[19]The contract could specify a transfer for any quantity above some lower limit and zero if less than that quantity is served, but the result is identical.

where $\hat{\theta}(\theta)$ is the firm's announcement strategy given the true demand state is θ, i.e. $\hat{\theta} : \Theta \to \Theta$. Again, applying the envelope theorem yields the condition

$$\frac{dU}{d\theta} = -\frac{\partial g}{\partial \theta}. \tag{A-14}$$

The necessary second order condition for maximization is

$$\left.\frac{\partial^2 U}{\partial \hat{\theta}^2}\right|_{\hat{\theta}=\theta} = \frac{d^2 p}{d\hat{\theta}^2}\hat{x}\big(p(\theta),\theta\big) + \left[\frac{dp}{d\hat{\theta}}\right]^2 \frac{\partial \hat{x}}{\partial p} + \left(\frac{\partial p}{\partial \hat{\theta}} - \frac{\partial^2 g}{\partial x^2}\frac{\partial \hat{x}}{\partial p}\frac{\partial p}{\partial \hat{\theta}}\right)\frac{\partial \hat{x}}{\partial p}\frac{\partial p}{\partial \hat{\theta}} + \frac{d^2 T}{d\hat{\theta}^2} \leq 0. \tag{A-15}$$

From (A-13), $\frac{d^2T}{d\theta^2}(\theta)$ satisfies the following equality

$$\frac{d^2 T}{d\theta^2} = -\frac{d^2 p}{d\hat{\theta}^2}\hat{x}\big(p(\theta),\theta\big) - \left[\frac{dp}{d\hat{\theta}}\right]^2 \frac{\partial \hat{x}}{\partial p} - \left(\frac{\partial p}{\partial \hat{\theta}} - \frac{\partial^2 g}{\partial x^2}\frac{\partial \hat{x}}{\partial p}\frac{\partial p}{\partial \hat{\theta}}\right)\frac{\partial \hat{x}}{\partial p}\frac{\partial p}{\partial \hat{\theta}} - \frac{\partial p}{\partial \theta}\frac{\partial \hat{x}}{\partial \theta}. \tag{A-16}$$

Plugging (A-16) into (A-15) yields

$$\left.\frac{\partial^2 U}{\partial \hat{\theta}^2}\right|_{\hat{\theta}=\theta} = -\frac{\partial p}{\partial \theta}\frac{\partial \hat{x}}{\partial \theta} \leq 0 \tag{A-17}$$

Because the firm's profit is concave in output $sign\big[\frac{d\hat{x}}{d\theta}\big] = sign\big[\Pi_{x\theta}\big]$, and $\Pi_{x\theta}$ is

$$\Pi_{x\theta} = -\frac{\partial^2 g}{\partial \hat{x}\partial \theta} \geq 0. \tag{A-18}$$

Thus $\frac{\partial^2 U}{\partial \hat{\theta}^2}(\theta,\theta) \leq 0$ requires $\frac{dp}{d\theta} \geq 0$.

To show sufficiency, start with

$$\frac{\partial U(\hat{\theta},\theta)}{\partial \hat{\theta}} = \frac{dp}{d\hat{\theta}}x + \frac{dT}{d\hat{\theta}}. \tag{A-19}$$

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

$$\frac{dT}{d\hat{\theta}} = -\frac{dp}{d\hat{\theta}}x\big(p(\hat{\theta}),\hat{\theta}\big). \tag{A-20}$$

Plugging (A-20) into (A-19) yields

$$\frac{\partial U(\hat{\theta},\theta)}{\partial \hat{\theta}} = \frac{dp}{d\hat{\theta}}\big[x\big(p(\hat{\theta}),\theta\big) - x\big(p(\hat{\theta}),\hat{\theta}\big)\big]$$

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

$$\frac{\partial U(\hat{\theta},\theta)}{\partial \hat{\theta}} = \frac{dp}{d\hat{\theta}}\frac{\partial x\big(p(\hat{\theta}),\tilde{\theta}\big)}{\partial \theta}(\theta - \hat{\theta}). \tag{A-21}$$

Because $dp/d\theta \geq 0$ and $\partial x/\partial\theta \geq 0$, (A-21) implies

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

Thus, $\hat{\theta} = \theta$ is a global maximizer.

PROOF OF PROPOSITION 3: The optimal unit price and transfer payments $\left\{p^{**}_{FP}(\theta), T^{**}_{FP}(\theta)\right\}_{\theta\in\Theta}$ were derived in the text. All that remains to be shown is that the unit price is increasing in the demand state to confirm that condition *(ii)* of lemma 2 is satisfied.

The unit price follows from (15) and the transfer payment is the transfer leaving the firm with the rents identified in (14). Taking the full derivative of p^{**}_{FP} yields the following derivative:

$$\frac{dp^{**}_{FP}}{d\theta} = \frac{V_{x\theta}(x,\theta) + \lambda\left(\frac{dH^{-1}}{d\theta}(\theta)g_{x\theta}(x,\theta) + H^{-1}(\theta)g_{x\theta\theta}(x,\theta)\right)}{(1+\lambda) - V_{xx}x_p - H^{-1}(\theta)g_{\theta xx}x_p} \tag{A-22}$$

The components of $dp^{**}_{FP}/d\theta$ have the following signs:

$$V_{x\theta}(x,\theta) \geq 0,$$
$$\frac{dH^{-1}}{d\theta}(\theta)g_{x\theta}(x,\theta) \geq 0,$$
$$H^{-1}(\theta)g_{x\theta\theta}(x,\theta)\big) \geq 0,$$
$$V_{xx}(x,\theta)\hat{x}_p \leq 0,$$
$$H^{-1}(\theta)g_{\theta xx}(x,\theta)\hat{x}_p \leq 0.$$

where the second result follows from the monotone hazard rate property. Therefore $dp^{**}_{FP}/d\theta \geq 0$ for all $\theta \in \Theta$.

The signs on the derivatives are all due to the assumptions made on the functions V and g. It should be noted that the signs on the derivatives are sufficient for $dp^{**}_{FP}/d\theta \geq 0$ but not necessary. When $\frac{dp}{d\theta} < 0$, then it is necessary for the contract to be non-linear in output. However, a contract which is non-linear in output will not be able to secure a more efficient outcome for the regulator as condition *(i)* of lemma 2 still identifies the amount of information rents the firm must obtain for incentive compatibility. See Rogerson (1987) for a thorough analysis describing when the optimal contract can be implemented with linear prices.

PROOF OF PROPOSITION 4: Proposition 4 can be proved with the following lemmas. First, letting γ denote the Lagrange multiplier, the first-order conditions for the firm's problem yield

$$\gamma = \frac{V_x(x, \theta)}{g_x(x, \theta) - p}, \tag{A-23a}$$

$$px - g(x, \theta) + T = 0, \tag{A-23b}$$

$$0 < \gamma \leq \infty. \tag{A-23c}$$

Because x must take a positive value, corner solutions are ignored. $\gamma > 0$ follows from the fact that $V_x > 0$ and $p \leq g_x(\hat{x}, \theta)$, where $\hat{x}$ is the maximand of (16). This is true because if, to the contrary, p exceeds the quality-adjusted marginal cost of producing the quantity $\hat{x}$, then the firm could induce a higher equilibrium quantity by increasing the quality of the service, all without losing profit. This relationship to price and quality-adjusted marginal cost is formally stated as lemma 4.

Lemma 4. *Let $\hat{x}$ solve the firm's optimization program (16), then $p \leq g_x(\hat{x}, \theta)$ for all feasible payment policies $\{p, T\}$.*

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 I 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

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

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

Lemma 5. *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 $\hat{x}$ to x^E.

The following lemmas characterize the first-best policy.

Lemma 6. *A policy inducing the first-best must include a positive lump-sum transfer for all quantities $0 < x^* < x^E$.*

Proof. Given a contract $\{p, T\}$, by plugging (A-23a) into (A-23b) allows the Lagrangian γ^* to be expressed as

$$\gamma^* = \frac{V_x(x,\theta)x}{g_x(x,\theta)x - g(x,\theta) + T}. \tag{A-24}$$

Because, $\gamma^* \geq 0$ by Lemma 4, it must be the case that $T \geq g(x,\theta) - g_x(x,\theta)x$. We can define the *quality adjusted* returns to scale as

$$s(x) = \frac{g_x(x,\theta)}{g(x,\theta)}. \tag{A-25}$$

The measure $s(x)$ has the usual interpretations, when $s(x) < 1$, the firm exhibits increasing returns to scale, and when $s(x) > 1$, the firm exhibits decreasing returns to scale. Using the quality adjusted returns to scale, it is clear that when $MC_{aq} < AC_{aq}$ any price less than marginal cost will result in a negative profit for the firm without a positive transfer.

Because $s(x^E) = 1$, a price equal to the quality-adjusted marginal cost will result in zero profits for the firm at x^E. Any increase in output will result in negative profits so the firm will choose x^E and no lump-sum transfer is needed. □

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, from lemma 4, 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 7. *First-best can be induced with only a unit payment, p, when $x^E \leq x^*$.*

Proof. Following from Lemma 6, when $s(x) \geq 1$ the price can be set equal to $AC_{aq}(x^*)$. Because $AC_{aq}(x^*) \leq MC_{aq}(x^*)$ the firm will choose the equilibrium quantity x^* and earn zero profits. □

Lemma 7 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 8. *First-best can be induced with only a lump-sum transfer, T, for any $x^* > 0$.*

Proof. Because $g_x > 0$ the firm's costs are increasing with x. Thus, the regulator need only set $T = g(x^*, \theta)$ to induce the firm to produce x^*. □

Combining lemmas 6 - 8 yields the proposition.

PROOF OF LEMMA 3: The necessity of the conditions were shown in the text. We will now show that the conditions are sufficient. Recall that

$$\left.\frac{d\hat{x}}{d\hat{\theta}}\right|_{\hat{\theta}=\theta} = \left.\frac{d\hat{x}}{dp}\frac{dp}{d\hat{\theta}}\right|_{\hat{\theta}=\theta} + \left.\frac{d\hat{x}}{dT}\frac{dT}{d\hat{\theta}}\right|_{\hat{\theta}=\theta} = 0, \tag{A-26}$$

implying

$$\frac{dT}{d\hat{\theta}} = -\frac{dp}{d\hat{\theta}} x\big(p(\hat{\theta}), T(\hat{\theta}), \hat{\theta}\big). \tag{A-27}$$

Plugging (A-27) into (A-26) yields

$$\frac{d\hat{x}}{d\hat{\theta}} = \frac{dp}{d\hat{\theta}} \left(\frac{d\hat{x}}{dp} - x\big(p(\hat{\theta}), T(\hat{\theta}), \hat{\theta}\big) \frac{d\hat{x}}{dT} \right). \tag{A-28}$$

Using the fact that $d\hat{x}\,/dp = -\Pi_p/\Pi_x$ and $d\hat{x}\,/dT = -\Pi_T/\Pi_x$ it is easy to show that

$$\frac{d\hat{x}}{dp} = \frac{d\hat{x}}{dT} x\big(p(\hat{\theta}), T(\hat{\theta}), \theta\big). \tag{A-29}$$

Plugging (A-29) into (A-28) yields

$$\frac{d\hat{x}}{d\hat{\theta}} = \frac{dp}{d\hat{\theta}} \frac{d\hat{x}}{dT} \big[x\big(p(\hat{\theta}), T(\hat{\theta}), \theta\big) - x\big(p(\hat{\theta}), T(\hat{\theta}), \hat{\theta}\big) \big] \tag{A-30}$$

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

$$\frac{dx(p(\hat{\theta}), T(\hat{\theta}), \theta)}{d\hat{\theta}} = \frac{dp}{d\hat{\theta}} \frac{d\hat{x}}{dT} \frac{\partial \hat{x}}{\partial \theta} (\theta - \hat{\theta}). \tag{A-31}$$

Because $dp\,/d\theta \geq 0$, $d\hat{x}\,/dT > 0$, and $\partial\hat{x}/\partial\theta \geq 0$, (A-31) implies

$$\frac{dx(\hat{\theta}, \theta)}{d\hat{\theta}} \geq 0 \text{ if } \hat{\theta} < \theta$$

$$\frac{dx(\hat{\theta}, \theta)}{d\hat{\theta}} \leq 0 \text{ if } \hat{\theta} > \theta$$

Thus, $\hat{\theta} = \theta$ is a global maximizer.

PROOF OF PROPOSITION 5: The conditions of Lemma 2 identify the necessary and sufficient conditions for an incentive compatible payment policy that must be satisfied by any set of payment rules: $\big\{p^{**}_{NP}(\theta), T^{**}_{NP}(\theta)\big\}_{\theta\in\Theta}$. The unit payment that induces the first-best level of output is derived by setting $d\hat{x}\,/d\theta = dx^*\,/d\theta$ and solving for p. By the integral form of the envelope theorem, we have $\hat{x}(p(\hat{\theta}), \theta) = x^*(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} (\partial x^*(\tilde{\theta})/\partial\theta) d\tilde{\theta}$, if and only if $\hat{x} = x^*$ at every $\theta \in \Theta$, thus a payment policy that induces $d\hat{x}\,/d\theta = dx^*\,/d\theta$ and sets $\hat{x}(\underline{\theta}) = x^*(\overline{\theta})$ induces the first-best output in every state.

From Proposition 4, the unit payment must always be less than or equal to the marginal cost at the induced quantity. Because $g_\theta(\cdot) \leq 0$, this condition requires that the first-best output be weakly decreasing with the cost-state, $dx^*\,/d\theta \leq 0$, otherwise, $p^{**}_{NP}(\theta) > g_x\big(\hat{x}(\theta), \theta\big)$ and the firm will choose a higher output. It must also be the case that $p(\theta) \neq g_x\big(\hat{x}(\theta), \theta\big)$ for all $\theta \in \Theta$, otherwise $d\hat{x}\,/d\theta$ is not bounded at some $\theta \in \Theta$, hence not absolutely continuous, and the envelope theorem cannot apply. The assumptions on the value and cost functions insure this cannot happen. First, the regulator's problem insures that $p^{**}_{NP}(\theta)$ is unique for every state θ; and second, the firm's problem insures that $\hat{x}$ is unique for every p. Therefore, for a given state the first-best quantity is unique and there does not exist any such $\theta' \in \Theta$ such that $\lim_{\theta \to \theta'} dx^*(\theta)/d\theta = \infty$.