Discussion Paper No. 1069

Choice of Treatment Intensities by A Nonprofit Hospital Under Prospective Pricing

by

William P. Rogerson*

First Version: Septenber 1993
Second Version: October 1993

*This paper was prepared for the conference "The Industrial Organization of Health Care," September 20-21, 1993. Research for this paper was supported by the Department of Veterans' Affairs. I would like to thank David Dranove, Kathleen Hagerty, Tom McGuire, Mike Riordan, Dan Spulber, and Will White for helpful discussions and comments.

1
Introduction

The last decade has witnessed a dramatic shift in the method used to determine hospitals' reimbursement for services they provide. Prior to 1984, most reimbursement was cost-based. In 1984, the federal government introduced a prospective payment system (PPS) for Medicare patients whereby hospitals received a fixed price for each payment depending on the diagnosis related group (DRG) of the patient. Since this time, many state governments have adopted similar systems for Medicaid patients and many private insurers have begun negotiating similar reimbursement systems for their patients. Thus a large fraction of hospital revenues are now determined by some type of prospective payment system.[1]

The primary motivation for this change was the hope that this system would reduce costs. The hoped-for cost reduction would occur for two quite distinct reasons. First, fixed prices would encourage hospitals to be efficient since they would become the residual claimant. Second, and perhaps most importantly, there was a feeling that demand for intensity of treatment was much higher than socially optimal due to the existence of insurance. By setting fixed prices for treating various DRGs, a payer (government or the private insurer) could indirectly control the intensity of treatment provided. In particular, by lowering prices (or by not allowing them to rise along with inflation over time) the payer could force hospitals to provide lower levels of treatment intensity (presumably, closer to the socially optimal level) and thus reduce health care costs.

To summarize the above, payers for health care now use price regulation of hospitals as a way of indirectly regulating the provision of treatment intensity. Most hospitals are nonprofits. Thus price regulation may not be needed to control profits and indirect control of treatment intensity might be viewed as the primary purpose of this price regulation.

For the case of a nonprofit hospital producing a single medical service, the correct course of action is clear. The payer should determine the level of intensity that it feels is appropriate and estimate the hospital's average cost of supplying this level of intensity. If price is set equal to this value of average cost, the hospital will be unable to raise treatment intensity above this level and break even. Assuming that the hospital is altruistic in the sense that it attempts to provide the best treatment possible subject to the constraint that it break even, the hospital will therefore choose intensity equal to the desired level.

In reality, of course, hospitals produce multiple services and thus the payer must set a price for each service. Given that the major role of price regulation is controlling intensity choice, and given that hospitals invariably produce multiple services, the question of how the simple one-product analysis generalizes to the case of multiple products is obviously important. The important questions are:

(i) How does a nonprofit hospital choose treatment intensities across products as a function of the price vector chosen by the payer?

(ii) Given this behavior, how should the payer optimally choose prices?

Surprisingly, there appears to be no existing formal analysis of these questions. The implicit assumption of policy analysis seems to be that the logic of the single product case can be applied on a product-by-product basis. That is, the answer to question (i) is that the hospital will choose treatment intensity to break even on a product-by-product basis. If this is true, then the answer to question (ii) is clear. For each product, the payer should decide on the desired treatment intensity, estimate the cost of supplying this intensity, and then set price equal to that cost. This theory of hospital decision making has some intuitive appeal. If the hospital is a nonprofit, it must break even overall. If the hospital is altruistic, in the sense that it agrees with the payer on the social value of increasing intensity across products, it may seem appealing to conclude that the hospital will "obey the command" of the payer and choose intensity on each product to break even.

This paper shows that, while perhaps superficially appealing, this theory is incorrect. The logically correct statement of an altruistic nonprofit hospital's decision problem is that it attempts to maximize the social value of health care (using the payer's definition of social value) subject to the constraint that it break even. It is shown that the solution to this problem generally does not involve the hospital breaking even on each product. In equilibrium, the ratio of price to marginal cost will vary across products inversely with the elasticity of demand with respect to treatment intensity. That is, products where demand is more responsive to increases in treatment intensity must exhibit lower ratios of price to marginal cost in equilibrium.

In this paper's model, the payer is generally able to induce the efficient or first best vector of treatment intensities. However, it does not do this by choosing prices such that the firm would break even on a product-by-product basis. Rather, it chooses prices so that, at the desired intensity vector, the ratio of price to marginal cost varies inversely with the intensity elasticity of demand. (The precise formula is given in the paper.) Generally speaking this will mean that the hospital earns positive profits on products with below average intensity elasticities and negative profits on products with above average intensity elasticities.

The intuition for this result depends on recognizing the nature of the "flaw" in the hospital's preferences. The hospital is altruistic in the sense that it evaluates the social value of health care using the payer's utility function over health care. However, this does not mean that the hospital and payer have the same preferences. Consider the following thought experiment. Suppose there are two different vectors of treatment intensities. One must be chosen. In either case the payer will reimburse the hospital for its costs so the hospital breaks even. How will the payer and hospital compare these alternatives? The payer will select the alternative that maximizes the social value of health care minus the cost of health care. The hospital, on the other hand, will select the alternative that maximizes the social value of health care. In particular, the cost of producing health care is of no direct relevance to the hospital. So long as it can induce the payer to pay, more is always better than less.

Thus the basic divergence in preferences between the hospital and payer is that the hospital only directly cares about the benefits of health care while the payer cares about the benefits minus costs. In particular, the hospital would always like the payer to spend more money on health care.

Now suppose that the payer wishes to induce a particular intensity vector and attempts to do this by setting prices so the firm would break even by choosing the desired vector. It turns out that, by increasing intensity on goods with high intensity elasticities and decreasing intensity on goods with low intensity elasticities, the hospital can simultaneously induce the payer to spend more and still break even. The hospital prefers this expenditure increase because it increases the social value of health care, even if it decreases the social value minus cost of health care.

Thus the hospital is biased towards increasing (decreasing) expenditures on high (low) elasticity products because this increases the payer's expenditures. Of course the hospital is also more predisposed to increase intensity on goods where the ratio of price to marginal cost is high (because profits earned on one good can be used to cover losses on other goods). By definition, in equilibrium the hospital must view the marginal value per dollar from increasing intensity to be the same across all products. Thus, products with high intensity elasticities (making the marginal value of intensity high) must be counterbalanced by low ratios of price to marginal cost (making the marginal value of intensity low).

In summary, this paper presents a theory of how a nonprofit altruistic hospital selects a vector of treatment intensities given a vector of prices and then shows how the payer should select prices in light of this theory. The payer is generally able to induce the hospital to select the efficient vector of intensities, but this is accomplished by choosing prices such that the ratio of price to marginal cost is higher on products whose demand is less responsive to intensity changes. This result is a byproduct of the fact that an altruistic hospital always prefers more health care to less so long as the payer can be induced to pay for it.

It is also shown that this paper's main result (that the ratio of price to marginal cost varies inversely with intensity elasticity across products) generalizes to a variety of different specifications of the hospital's objective function.[2] Two major alternate specifications are considered where precisely the same result is shown to hold. Under the first, the hospital is assumed to be less altruistic than in the main model. It is assumed that the hospital's goal is to maximize its size (as measured by revenues) subject to a break-even constraint. Under the second, the hospital is assumed to be less altruistic, still. It is assumed that the hospital values both profit and revenues, and attempts to maximize a utility function defined over both these variables. This second alternative formalizes the idea that nonprofits may value profit to the extent that it can be converted into managerial perks, used to fund projects of particular value to managers, etc. Under both of these alternate specifications, the result that price varies inversely with intensity elasticity across products continues to hold true.

Finally, the case where the hospital has direct preferences over the health care it provides, but these preferences are different than the payer's is considered. (Recall that, in the basic model the hospital is assumed to be altruistic in the sense that its preferences over health care are the same as the payer's.) This case is meant to capture the idea that the hospital may care directly about some aspects of health care such as technological sophistication more than does the payer. It is shown that the ratio of price to marginal cost is now affected by two factors. Just as in the basic model, the ratio varies inversely with intensity elasticity. However, the ratio also varies inversely with the hospital's own assessment of the marginal value of each service it provides, i.e., the ratio of price to marginal cost is lower on services that the hospital is particularly anxious to provide more of. Thus the effect identified in the basic model continues to hold, although it is now only one of two effects determining the ratio of price to marginal cost.

A relatively large number of papers in industrial organization, regulation, and health care have considered the subject of the effects of price regulation on quality choice. Most of these papers consider the case of a firm whose goal is profit maximization.[3] One exception is Hodgkin and McGuire (1993) which assumes that the hospital cares about both profit and intensity. However, they only consider the single product case. Harris's (1979) analysis of how an altruistic hospital sets prices across products when consumers are differentially insured is similar in spirit to this paper's analysis. However, Harris does not consider intensity as a choice variable for the hospital. Two papers by Rogerson (1990, 1991) build models of bureaucratic decision making where bureaucrats are altruistic in the same sense as the hospital is in this paper, i.e., their goal is to maximize the social benefit of their activity and they do not directly care about the cost. Just in this paper, this means that bureaucrats choose actions in an attempt to increase expenditures on their activity. However, the types of decisions modelled and the nature of the distortions which result are very different in these papers.

Section 2 presents the model. Section 3 presents preliminary analysis and section 4 presents the main results concerning the ratio of price to marginal cost across products. Section 5 considers the issue of whether the results of section 4 hold true when accounting costs are substituted for marginal costs. Section 6 generalizes the basic model to allow for expenditure caps and suggests that expenditure caps may play a useful role in a prospective pricing system. Section 7 generalizes the model to allow for a variety of other specifications of the hospital's objective function. Finally, section 8 discusses the results.

2
The Model

Part A will describe the utility and cost functions of the various agents in the model. Parts B and C will define, respectively, equilibrium and efficient outcomes. Then finally part D will introduce three assumptions that play a role in guaranteeing the existence of equilibrium and efficient outcomes.

A. Utility Functions and Cost Functions

Suppose that there is a single payer and a single hospital. Thus issues of strategic interaction among multiple payers or competition between multiple hospitals are abstracted away from.[4]

Suppose that there are n_d different diagnoses or illnesses that the hospital treats and n_g different groups of patients. In a larger model the payer would be modelled as choosing how to define diagnoses and groups of patients. In this model, these will be taken as given.

Define a "product" to be the treatment of a particular diagnoses for a patient of a particular group. Let $n = n_d n_g$ denote the number of products the hospital produces and let $i \in \{1,\ldots, n\}$ denote an index for the products.

Let q_i denote the intensity of treatment for product i where q_i is a non-negative real number. Intensity of treatment can be increased by increasing the number of tests, using more advanced technology, increasing the length of stay, etc. Let $q = (q_1,\ldots, q_n)$ denote the vector of intensities for all products. Let x_i denote the quantity supplied of product i where quantity is measured in discharges.[5] Let $x = (x_1,\ldots, x_n)$ denote the vector of quantities.

The vector (x,q) will be referred to as a health care outcome since it completely describes the health care produced.

Assume that demand for product i is determined by treatment intensity according to the function

$$x_i = x_i(q_i) \tag{2.1}$$

Let x(q) denote the vector $(x_1(q_1),\ldots, x_n(q_n))$. Patients are assumed to be fully insured so that price does not affect demand.[6] Also, the products' demands are independent in the sense that demand for product i is only affected by product i's treatment intensity. Assume that *(a.1)* $x_i(q_i)$ is a real-valued function mapping $[0,\infty)$ onto $[0,\infty)$. It is twice continuously differentiable, and strictly increasing.

Let $q_i(x_i)$ denote the inverse function and let $q(x) = (q_1(x),\ldots,q_n(x))$ denote the vector of intensity functions. Obviously, $q_i(x_i)$ satisfies the properties in (a.1), as well.

It will be assumed that the hospital chooses treatment intensity, q, but then must supply enough discharges to meet all demand thus generated, x(q).[7]

The payer values the benefits that patients receive from health care according to the utility function U(x,q). The payer's goal is to maximize the benefits of health care minus its payments to the hospital. The hospital is assumed to be a nonprofit and thus is constrained to earn nonpositive profit. In order to survive, it must earn non-negative profit. Thus the hospital must earn zero profit. It is assumed that the hospital is perfectly altruistic in the sense that its goal is to maximize patients' benefits of health care as viewed by the payer. Thus the hospital's goal is to maximize U(x,q) subject to the constraint that it earn zero profit.

Let V(x) denote the utility which occurs if x is the quantity and q is chosen to induce x,i.e.,

$$V(x) = U\big(x, q(x)\big) \quad . \tag{2.2}$$

Since V will be the relevant utility function for most of the analysis of this paper, it is most straightforward to state the conditions that preferences are assumed to satisfy in terms of V.

Assume that

(a.2) V is a real-valued twice continuously differentiable function defined over $(0,\infty)$. It is strictly increasing in each argument and strictly concave.

(a.3) $\lim_{x_i \to 0} V(x) = -\infty$.

Assumption (a.2) simply states that V is a well-behaved utility function. Assumption (a.3) guarantees that solutions to various maximization programs will never be corner solutions so attention can be restricted to values of $x > 0$.[8]

Let $F(x,q)$ denote the cost of producing (x,q). Define $G(x)$ to be the cost of producing x given that treatment intensity is adjusted to cause patients to demand x. It is formally defined by

$$G(x) = F(x, q(x)) \tag{2.3}$$

and will be referred to as the intensity adjusted cost function. The function G(x) will be the cost function used in most of the analysis. Therefore it is most convenient to directly introduce most of the required assumptions about cost functions in terms of G. Assume that

(a.4) G is a real-valued function defined over $[0,\infty)^n$ such that $G(0) = 0$ and $G(x) > 0$ for every $x \neq 0$. It is twice continuously differentiable, strictly increasing, and strictly convex over $(0,\infty)^n$.

Assumption (a.4) states that G exhibits smooth, strictly increasing, strictly convex variable costs. However, G may exhibit fixed costs and thus jump discontinuously as any x_i is increased from 0 to a small positive amount.

It will also be useful to formally make one very natural assumption about the cost function F which is not necessarily implied by (a.4). Assume that

(a.5) F is a real-valued function defined over $[0,\infty)^{2n}$ which is twice continuously differentiable and strictly increasing over $(0,\infty)^{2n}$.

In particular, costs increase in both x_i and q_i, which is natural to expect.

It is important to note that the intensity adjusted cost function, G(x), is an unusual cost function and is not what economists would normally refer to as the hospital's cost function. The function G(x) gives the cost of producing the quantity vector x given that treatment intensity is adjusted to induce patients to demand x. Economists would normally calculate a cost function for some fixed treatment intensity vector. A useful way to think of this distinction is to consider the formula for intensity adjusted marginal cost,[9]

$$G_i(x) = \frac{\partial F}{\partial x_i}(x, q(x)) + \frac{\partial F}{\partial q_i}(x, q(x))\, q_i'(x) \quad . \tag{2.4}$$

The first term on the RHS of (2.4), given by $\partial F/\partial x_i$, is the marginal cost of producing more units of x_i and will thus be referred to as the "marginal cost of production." The second term on the RHS of (2.4), given by $(\partial F/\partial q_i)q_i'$, is the marginal cost of increasing treatment intensity to induce patients to demand the extra units of x_i that are produced. Therefore, it will be referred to as the "marginal cost of demand inducement." The intensity adjusted marginal cost, G_i, is the sum of these two terms. By assumptions (a.1) and (a.5) both of these terms are positive. In particular, this means that the intensity adjusted marginal cost is strictly greater than the marginal cost of production.

Since economists would normally hold intensity fixed when they calculate the cost function for a hospital, they would calculate a value of marginal cost equal to marginal production cost. The intensity adjusted marginal cost is strictly greater than this because it includes, additionally, the cost of increasing intensity sufficiently to induce patients to demand the higher output.

A special case that will be referred to in the discussion of the results occurs when there are constant returns to scale and production of each good is separable. This occurs when F is given by

$$F(x, q) = \sum_{i=1}^{n} c_i(q_i) x_i \quad . \tag{2.5}$$

In order to guarantee that (a.4)–(a.5) are true it is sufficient to assume that[10]

(b.1) $c_i(q_i)$ is non-negative, twice continuously differentiable, strictly increasing and strictly convex over $[0, \infty)$.

(b.2) $q_i(x)$ is weakly convex over $[0, \infty)$.

The results of this paper are particularly striking and simple for this special case because average cost equals marginal cost. Thus marginal cost pricing allows the firm to break even. The result of this paper is that marginal cost pricing never occurs in equilibrium.

It will be useful to create a measure of the elasticity of a good's demand with respect to expenditures on treatment intensity. Let $\eta_i(x)$ denote this measure which is defined by

$$(2.6) \qquad \eta_i(x) = \frac{x_i'\left(q_i(x)\right)/x_i}{\frac{\partial F}{\partial q_i}\left(x,q(x)\right)/\frac{\partial F}{\partial x_i}\left(x,q(x)\right)x_i} \quad .$$

The numerator of (2.6) is the percentage change in demand for good i when q_i is increased. This is straightforward. The denominator of (2.6) can be interpreted as the percentage change in the cost of producing good i when q_i is increased. This interpretation is slightly more involved. The term $\partial F/\partial q_i$ is clearly the change in cost. In order to calculate the percentage change in cost, we must therefore divide by the total cost of producing good i. The problem is that, when costs are not separable, there is no unambiguous method of defining this cost. Therefore, the term $(\partial F/\partial x_i)x_i$ has been chosen as the estimate of the cost of producing good i. For the special linear, separable case of (2.5), it is clear that this choice is the unambiguous best one. In this case

$$(2.7) \qquad \frac{\partial F}{\partial x_i}x_i = c_i(q_i)x_i$$

which is clearly the cost of producing good i. In this case the elasticity is given by

$$(2.8) \qquad \eta_i(x) = \frac{x_i'(q_i)\,/\,x_i}{c_i'(q_i)\,/\,c_i(q_i)}$$

which is the elasticity of demand for good i with respect to intensity increasing expenditures which increase the unit cost of good i. In the nonseparable case there is no unambiguous correct choice and (2.6) can be interpreted as a reasonable definition for a measure of the elasticity of demand with respect to quality expenditures.

B. Equilibrium

The structure of the strategic interaction between the hospital and the payer can now be described. It is assumed that the payer and hospital both have symmetric and complete knowledge of all aspects of the model such as demand functions, cost functions, one another's preferences, etc. However, it is assumed that intensity cannot be contracted upon. The motivation for this assumption is that aspects of treatment intensity are often difficult to objectively specify and measure. As a consequence, the payer can only indirectly influence the hospital's intensity choice through its choice of prices. Therefore, the sequence of moves is as follows. First, the payer chooses a price vector p $= (p_1, \ldots, p_n)$. Given this price vector, the hospital then chooses an intensity vector, q, so as to maximize V(q,x(q)) subject to the constraint that it break even. Recall that a health care vector is as an ordered pair (x,q). However, since patients always choose quantity according to the demand function x(q), it is sufficient to describe an outcome by only specifying one of the two vectors, x and q. It is perhaps most natural to view q as the outcome description, since hospitals view themselves as choosing q. However, logically, one can also think of the hospital as choosing x (and then choosing q so that x is chosen). Although this latter viewpoint seems more convoluted, it turns out to be useful for generating clear, simple intuitions for the main results. Therefore, this paper will view x as the variable describing outcomes.

Formally, x will be said to be induced by p if it solves the program

$$\text{(2.9)} \qquad \underset{x>0}{\text{Maximize}} \quad V(x)$$

(2.10) subject to $\sum_{i=1}^{n} p_i x_i - G(x) = 0$

Let P(x) denote the (possibly empty) set of prices that induce x. Let I denote the set of outputs which are inducible in the sense that a price exists which induces them.

(2.11) $I = \{x : P(x) \neq \emptyset\}$.

An equilibrium can now be described. The payer can be viewed as choosing an (x,p) pair to maximize his utility of health care minus his payments for health care subject to the constraint that the hospital will choose x given p. Formally, then, (x,p) is an equilibrium if it solves the following program.

(2.12) $\underset{x>0,p}{\text{Maximize}} \quad V(x) - \sum_{i=1}^{n} p_i x_i$

(2.13) subject to $p \in P(x)$

However, by (2.10), if p $\in$ P(x) then the hospital earns zero profit, so the payer's payments to the hospital equal the hospital's costs. This means that the hospital's costs can be substituted for the payer's payments in the objective function (2.12). This yields an extremely simple characterization of an equilibrium.

A pair (x,p) is an equilibrium if and only if x solves the program

(2.14) $\underset{x>0}{\text{Maximize}} \quad V(x) - G(x)$

(2.15) subject to $x \in I$

and p satisfies

(2.16) $p \in P(I)$

That is, the payer chooses an optimal output by first finding the x in I that maximizes his utility minus the cost of production. Then any $p \in P(x)$ creates an equilibrium (x,p) pair.

C. Efficiency

For some expenditure level g, a vector x will be said to be efficient given the expenditure level g if it solves

(2.17) $$\underset{x>0}{\text{Maximize}} \quad V(x)$$

(2.18) $$\text{subject to} \quad G(x) \leq g$$

A quantity vector x will simply be called efficient if it maximizes benefits minus costs of health care, i.e., if it solves

(2.19) $$\underset{x>0}{\text{Maximize}} \quad V(x) - G(x)$$

D. Existence Assumptions

Three more assumptions will now be introduced which can all be interpreted as stating that "costs eventually begin to rise rapidly." They are all relatively natural, and their purpose is simply to guarantee existence to various optimization problems.

(a.6) For every $p \in R^n$, the set

$$\left\{ x : x \geq 0 \text{ and } \sum_{i=1}^{n} p_i x_i - G(x) \geq 0 \right\}$$

is bounded.

This states that the set of outputs that generate non-negative profit is always bounded, i.e., the firm cannot increase output infinitely and continue to earn positive profit.

Second, assume that

(a.7) For every $k \in R$, the set

$$\{x : x \geq 0 \text{ and } V(x) - G(x) \geq k\}$$

is bounded.

This assumption states that the outputs generating surplus greater than or equal to any given level are bounded.

The third assumption requires the introduction of a new concept. Define S(x) to be the elasticity of total cost with respect to a small proportionate change in all outputs. Formally,

$$S(x) = \left[\frac{d}{d\theta} G(\theta x_1, \ldots, \theta x_n)\right] \frac{\theta}{G(\theta x)}\bigg|_{\theta=1} \tag{2.20}$$

which can be rewritten as

$$S(x) = \frac{\sum_{i=1}^{n} G_i(x) x_i}{G(x)} \tag{2.21}$$

The function S(x) is normally interpreted as a measure of returns to scale at x.[11] If S(x) < (=,>)1, these are said to be increasing (constant, decreasing) returns to scale in the sense that a proportionate increase in all outputs causes a less than (equal, more than) proportionate increase in costs. Thus, there is a sense in which average costs drop (stay constant, increase).

The third assumption can now be stated. It will be assumed that there is at least one x > 0 such that G exhibits constant or decreasing returns to scale.

(a.8) There exists an x > 0 such that $S(x) \geq 1$.

This is, of course, an extremely minimal assumption. In fact it seems plausible that G(x) will generally exhibit decreasing returns to scale over large portions of the relevant range of production for most hospitals. Recall that G(x) includes both the cost of production and the cost of increasing intensity to induce patients to demand that quantity. The cost function normally considered by economists holds intensity fixed. This means that marginal cost is higher under G(x) than the cost function normally considered by economists. Therefore, G(x) will always exhibit lower returns to scale than will the cost function typically considered by economists. In particular, then, a sufficient condition for G to exhibit decreasing returns is that the normal constant intensity cost function exhibit decreasing or constant returns. There is some evidence that hospitals' constant intensity cost functions exhibit relatively constant returns to scale over large output ranges.[12] Thus it seems plausible that G(x) may generally exhibit decreasing returns to scale over a broad range of outputs.

It is straightforward to formalize this point. Let N(x) denote the returns to scale measure at x holding intensity fixed at q(x). It is given by

$$(2.22) \qquad N(x) = \frac{\sum_{i=1}^{n} \left\{ \frac{\partial F}{\partial x_i}\big(x, q(x)\big) \right\} x_i}{F\big(x, q(x)\big)}$$

Of course (2.21) can be rewritten as

$$(2.23) \qquad S(x) = \frac{\sum_{i=1}^{n} \left\{ \frac{\partial F}{\partial x_i}\big(x, q(x)\big) + \frac{\partial F}{\partial q_i}\big(x, q(x)\big) q_i'(x) \right\} x_i}{F\big(x, q(x)\big)}$$

These expressions are the same except for the marginal cost terms in brackets. For N(x), the marginal production cost is used. For S(x), the intensity adjusted marginal cost (i.e., the production marginal cost plus the demand inducement marginal cost) is used. Therefore, S(x) is strictly greater than N(x). In particular, S(x) will be greater than 1 if N(x) is greater than or equal to 1.

The special constant returns to scale cost function in (2.5) illustrates this idea. For this function, there are globally constant returns to scale holding q constant. Thus N(x) = 1. Since S(x) is strictly greater than N(x), this implies that S(x) is globally strictly greater than 1.

3
Preliminary Analysis

Part A of this section will characterize the correspondence P(x) and thus also characterize the set of inducible outputs, I. Then part B will use these results to characterize the equilibrium. Section 4 will go on to analyze the nature of price cost margins in equilibrium.

A. Characterization of P(x)

The program (2.9)-(2.10) defining P(x) uses an equality constraint for profit. This is because the firm must earn non-negative profit to survive and nonpositive profit by the legal requirement that it be a nonprofit. If the program were relaxed to allow any non-negative profit, it would become

$$\text{(3.1)} \qquad \underset{x>0}{\text{Maximize}} \quad V(x)$$

$$\text{(3.2)} \qquad \text{subject to} \quad \sum_{i=1}^{n} p_i x_i - G(x) \geq 0$$

Lemma 3.1 states that this relaxed program is equivalent to the unrelaxed program.[13]

Lemma 3.1 Fix any p ∈ R^n. Then x satisfies (2.9)-(2.10) if and only if x satisfies (3.1)-(3.2).

This result is very intuitive. If profits are positive, the hospital could always increase its utility from health care by supplying more health care. For the remainder of this paper program (3.1)-(3.2) will be viewed as the definition of x being induced by p.

The first order conditions for program (3.1)-(3.2) are

$$\text{(3.3)} \qquad \frac{V_i(x)}{G_i(x) - p_i} = \delta \qquad \text{for i} = 1,\ldots, \text{n}$$

$$\text{(3.4)} \qquad \sum_{i=1}^{n} p_i x_i - G(x) = 0$$

$$\text{(3.5)} \qquad 0 < \delta \leq \infty$$

Four remarks should be noted about these first order conditions. First, since x must be positive by (a.3), there is no need to consider corner solutions. Second, since lemma 3.1 establishes that the break-even constraint will be satisfied with equality, (3.4) is written with equality. Third, since $V_i > 0$ by (a.2), δ cannot be zero so (3.5) is written to reflect this. Fourth, and most important, δ is allowed to assume the value ∞. If $\delta = \infty$, conditions (3.3)-(3.4) can be written

$$\text{(3.6)} \qquad G_i(x) - p_i = 0 \qquad \text{for i} = 1,\ldots, \text{n}$$

$$\text{(3.7)} \qquad \sum_{i=1}^{n} p_i x_i - G(x) = 0$$

Equation (3.6) states the first order conditions for unconstrained profit maximization. Equation (3.7) states that profits are zero at this point. Thus (3.6)-(3.7) are the first order conditions corresponding to the case where there is a single $x > 0$ satisfying the constraint (3.2) and thus this single element is the solution to the program. Rather than write this as a separate case it is more convenient to simply let δ assume the value ∞.

It will be convenient to introduce notation to capture this distinction. If there exists an $x > 0$ such that

$$(3.8) \qquad \sum_{i=1}^{n} p_i x_i - G(x) > 0 \ ,$$

then p will be said to permit positive profit. If there does not exist an $x > 0$ such that (3.8) is true but there does exist an x such that (3.8) is satisfied with equality, it will be said that p permits zero but not positive profit.

Enough regularity has been assumed to guarantee that the first order conditions characterize the solutions to (3.1)-(3.2). This is stated as lemma 3.2.

Lemma 3.2 For any (x,p), where $x > 0$, x is induced by p if and only if there exists a δ such that (3.3)-(3.5) are satisfied. If $\delta < \infty$, then p permits positive profit. If $\delta = \infty$ then p permits zero but not positive profit.

For any $x > 0$, define $p^*(x) = (p_1^*(x),\ldots,p_n^*(x))$ and $\delta^*(x)$ as follows.

$$(3.9) \qquad p_i^*(x) = G_i(x) - \frac{V_i(x)}{\delta^*(x)}$$

$$(3.10) \qquad \delta^*(x) = \frac{\sum_{i=1}^{n} V_i(x) x_i}{\sum_{i=1}^{n} G_i(x_i) x_i - G(x)}$$

It is straightforward to invert (3.4)-(3.5) to show that for any $x > 0$, $p_i^*(x)$ and $\delta^*(x)$ are the unique values of p and δ such that (x,p,δ) satisfies (3.4)-(3.5). This is stated as lemma 3.3.

Lemma 3.3 For any $x > 0$, $p^*(x)$ and $\delta^*(x)$ are the unique values of p and δ such that (x,p,δ) satisfies (3.3)-(3.4).

If follows immediately that for any $x > 0$, that x is inducible if and only if $\delta^*(x) > 0$. If x is inducible, then $p^*(x)$ is the unique price that induces it. The price $p^*(x)$ permits positive profit if $\delta^*(x) < \infty$ and permits zero but not positive profit if $\delta^*(x) = \infty$. Equation (3.10) determines $\delta^*(x)$. The numerator of (3.10) is always positive. Therefore $\delta^*(x)$ is negative, positive, or infinite, depending upon whether the denominator,

$$\sum_{i=1}^{n} G_i(x_i)x_i - G(x) \quad , \tag{3.11}$$

is negative, positive, or zero. By inspecting the definition of the returns to scale measure, S(x), in (2.21), it is clear that (3.11) is negative, positive, or zero, depending upon whether S(x) is less than, greater than, or equal to 1.

Therefore, x is inducible if and only if the cost function G exhibits constant or decreasing returns to scale at x. If x is inducible, $p^*(x)$ is the unique price which induces it. If G(x) exhibits constant returns to scale at x, then $p^*(x)$ permits zero but not positive profit. If G exhibits decreasing returns to scale at x then $p^*(x)$ permits positive profit. This conclusion is formally stated as proposition 3.1.

Proposition 3.1

Suppose that $x > 0$. Then x is inducible if and only if $S(x) \geq 1$. If x is inducible, $p^*(x)$ is the unique price which induces it. If $S(x) = 1$, then $p^*(x)$ permits zero but not positive profit. If $S(x) > 1$, then $p^*(x)$ permits positive profit.

Proposition 3.1 states that x is inducible if and only if the cost function G exhibits constant or decreasing returns to scale at x. It is intuitively clear that decreasing or constant returns to scale are a necessary condition for inducibility. Suppose, for contradiction, that x is induced by some price p and that there are increasing returns to scale at x. Now consider a proportionate increase in all outputs. The firm's revenue rises proportionately since all prices are constant. By assumption, the firm's costs rise less than proportionately. Therefore, the new, increased, output vector is feasible because profits are positive. However, since all outputs are higher, the hospital prefers this vector.

To summarize this logic, an altruistic hospital can never be induced to choose any output where there are increasing returns to scale in the cost function, G(x). This is because so long as the hospital's is breaking even at x, it can continue to break even by proportionately increasing x. Thus a necessary condition for x to be inducible is that G(x) exhibit constant or decreasing returns to scale. Proposition 3.1 shows that this condition is, in fact, also sufficient for inducibility and shows that $p^*(x)$ is the unique price that induces x when it is inducible. Furthermore, $p^*(x)$ permits positive profit when $S(x) > 1$ and permits zero but not positive profit when $S(x) = 1$.

B. Equilibrium

It is straightforward to show that a unique efficient output exists. Lemma 3.4 states this result.

Lemma 3.4 There exists a unique x which is efficient.

Let x^E denote the unique efficient output.

By proposition 3.1, the definition of an equilibrium in (2.14)-(2.16) can be rewritten as follows. A pair (x,p) is an equilibrium if x solves the program

$$\text{(3.12)} \qquad \underset{x>0}{\text{Maximize}} \quad V(x) - G(x)$$

$$\text{(3.13)} \qquad \text{subject to} \quad S(x) \geq 1 \ ,$$

and p satisfies

$$\text{(3.14)} \qquad p = p^*(x) \ .$$

Sufficient regularity assumptions have been made to guarantee that program (3.12)-(3.13) has a solution. Thus an equilibrium always exists. This is stated as lemma 3.5.

Lemma 3.5 There exists a solution to program (3.12)-(3.13).

The nature of the solution will now be characterized. Recall that x is efficient if it solves

(3.15) $$\underset{x>0}{\text{Maximize}} \quad V(x) - G(x)$$

Therefore under both program (3.12)-(3.13) and (3.15), x is chosen to maximize the same objective function. The difference is that under the equilibrium program, the constraint is added that $S(x) \geq 1$.

Recall that x^E is defined to be the unique efficient point. If follows immediately from the above that if G exhibits constant or decreasing returns to scale at x^E, that x^E is the equilibrium quantity. This is stated as lemma 3.6.

Lemma 3.6 Suppose that $S(x^E) \geq 1$. Then $(x^E, p^*(x^E))$ is the unique equilibrium.

If $S(x^E) < 1$, then x^E cannot be chosen. Since the objective function is concave, it is immediate that the constraint (3.13) must hold with equality. This conclusion is stated as lemma 3.7.

Lemma 3.7 Suppose that $S(x^E) < 1$. Then if (x,p) is an equilibrium, $S(x) = 1$.

Proposition 3.2 summarizes the conclusions of this part.

Proposition 3.2

(i) There exists a unique efficient output denoted by x^E.

(ii) If $S(x^E) \geq 1$, the unique equilibrium output is x^E.

(iii) If $S(x^E) < 1$, at least one equilibrium output exists.
Every equilibrium satisfies $S(x) = 1$.

4
Price Cost Ratios

This section will characterize how the ratio of price to marginal cost varies across products in equilibrium. In the model of this paper, it is possible to calculate these ratios using either the intensity adjusted marginal cost, G_i, or the marginal cost of production, $\partial F/\partial x_i$. In real practice, economists calculate cost functions for hospitals holding intensity fixed. Therefore, the value of marginal cost that they would normally calculate is the marginal cost of production. Therefore, in order to characterize the nature of the ratio price to marginal cost that would normally be calculated by economists, one should use the marginal cost of production. However, it turns out that considering the ratio of price to intensity adjusted marginal cost sheds some extra light on the structure of the model. Therefore, this will also be considered. Part A will characterize price/marginal cost ratios using intensity adjusted marginal cost. Then part B will do the same using the marginal cost of production. Part C will describe the implications of this result for the separable constant returns to scale case. Finally, part D will discuss the connection of the results to Ramsey pricing.

A. Intensity Adjusted Marginal Cost

Recall that x is inducible if and only if $S(x) \geq 1$. When x is inducible, $p^*(x)$, defined by (3.9) is the unique price that induces x. Dividing both sides of (3.9) by $G_i(x)$ yields the following result.

$$(4.1) \qquad \frac{p_i}{G_i(x)} = 1 - \frac{V_i(x)}{G_i(x)} \cdot \frac{1}{\delta^*(x)}$$

where $\delta^*(x)$ is defined by (3.10).

If $S(x) = 1$, then $\delta^*(x) = \infty$ and p_i/G_i is constant and equal to 1 for every i. This result is stated as lemma 4.2.

Lemma 4.2 Suppose that p induces x and $S(x) = 1$. Then

$$\frac{p_i}{G_i(x)} = 1 \quad . \tag{4.2}$$

This result is very intuitive. If $S(x) = 1$, setting $p_i = G_i(x)$ for every i means that x maximizes the firm's profit and the profit at x equals zero. Thus x is the unique element satisfying the break-even constraint so the firm chooses it.

The situation where $S(x) > 1$ is more complex. In this case, $\delta^*(x) \in (0,\infty)$. Therefore by (4.1), p_i/G_i varies inversely with the ratio V_i/G_i. This result is very intuitive. A higher value of V_i/G_i means that expanding output of product i is more desirable in the sense that it yields more utility per dollar. If one attempts to induce the hospital to choose an output vector where it is more desirable to increase some outputs than others, the ratio of price to marginal cost must be chosen to be lower on the more desirable products. (Note that even though the hospital is nonprofit, the ratio of price to marginal cost still affects its output decision. All other things equal, increasing the p_i/G_i ratio makes increasing the output of product i more attractive because the cost savings can be used to produce more output of some other good and thus increase the utility from health care.)

Therefore, when $S(x) > 1$, the fact that p induces x does not necessarily imply that p_i/G_i is constant. This is only true if the utility per dollar from increased output, V_i/G_i, is constant across products. However, by proposition 3.2 this is true if x is an equilibrium. By proposition 3.2, if x is an equilibrium and $S(x) > 1$ then x equals x^E, the efficient output. Of course the first order condition for x^E is

$$V_i(x) = G_i(x) \quad . \tag{4.3}$$

Substitution of (4.3) into (4.1) therefore yields the following result.

Lemma 4.3 Suppose that (x,p) is an equilibrium and $S(x) > 1$. Then

$$\frac{p_i}{G_i(x)} = \frac{1}{S(x)} \tag{4.4}$$

In particular, the ratio $p_i/G_i(x)$ is constant across products. The intuition for this result is illustrated in figure 4.1 for the case of two products. The efficient output is labelled x^E. The iso-utility curve passing through x^E is labelled V and the iso-cost curve is labelled G. Since x^E is efficient, the two curves are tangent at x^E. The slope of the iso-cost curve at x^E is

$$\frac{dx_2}{dx_1} = \frac{-G_1(x^E)}{G_2(x^E)} \tag{4.5}$$

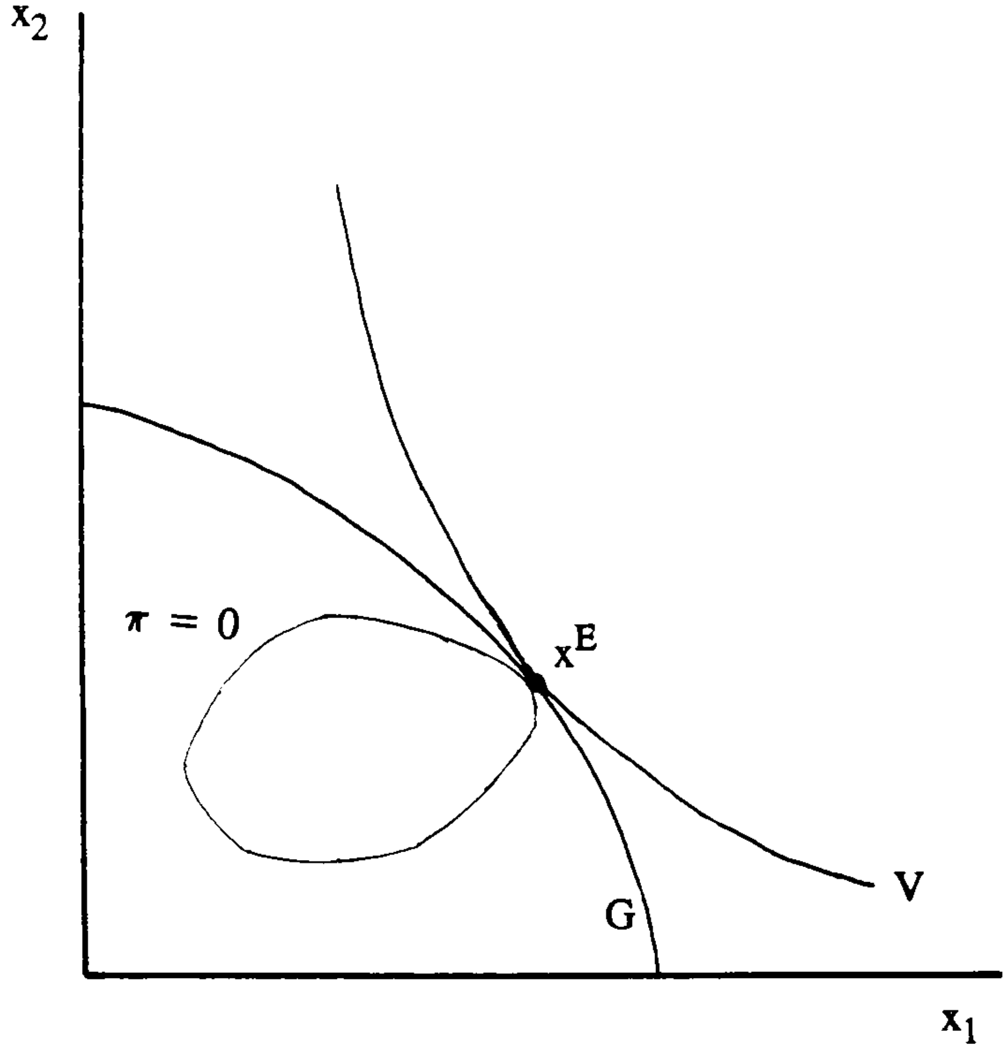
In order to choose a p which induces x^E, we must choose a p such that the iso-profit line for the zero profits level passes through x^E and is tangent to V. This is drawn as $\pi = 0$ in the figure. Then when the hospital chooses an x to maximize V subject to $\pi \geq 0$ it will clearly choose x^E. The slope of the iso-profit line at x^E is given by

$$\frac{dx_2}{dx_1} = \frac{-\left(p_1 - G_1(x^E)\right)}{p_2 - G_2(x^E)} \tag{4.6}$$

By comparing (4.5) and (4.6), it is clear that prices must be chosen proportionate to marginal costs in order for the iso-profit line to have the same slope as the iso-cost line.

It should also be noted that lemma 4.3 implies that in an equilibrium where $S(x) > 1$, price will always be strictly less than intensity adjusted marginal cost. This is simply a consequence of the fact that the firm earns zero profit in equilibrium. If there are decreasing returns to scale and prices

Figure 4.1
Equilibrium When $S(x^E) > 1$



are proportionate to marginal cost, then prices must be strictly less than marginal cost for the firm to earn zero profit.

Lemma 4.2 implies that p_i/G_i is constant when p induces x and S(x) = 0. In particular, then, p_i/G_i is constant when (x,p) is an equilibrium. Lemma 4.3 implies that p_i/G_i is constant when S(x) > 0 and (x,p) is an equilibrium. Therefore, lemmas 4.2 and 4.3 together imply that p_i/G_i is constant in equilibrium. This is the major result of this section and will be stated as proposition 4.1.

Proposition 4.1

Suppose that (x,p) is an equilibrium. Then

$$\frac{p_i}{G_i(x)} = \frac{p_j}{G_j(x)} \qquad \text{for every i,j.} \tag{4.7}$$

Of course, in equilibrium the firm also earns zero profit. That is, (x,p) satisfies

$$\sum_{i=1}^{n} p_i x_i - G(x) = 0 \quad . \tag{4.8}$$

Equations (4.7) and (4.8) are equivalent to

$$\frac{p_i}{G_i(x)} = \frac{1}{S(x)} \qquad \text{for every i.} \tag{4.9}$$

That is, when p_i/G_i is constant and profit is zero, p_i/G_i must equal 1/S(x). Therefore, corollary 4.1, below, is equivalent to proposition 4.1.

Corollary 4.1

Suppose that (x,p) is an equilibrium. Then (4.9) is true.

Although proposition 4.1 and corollary 4.1 are equivalent, it will be convenient to view proposition 4.1 as the statement of the main result. This focuses attention on the result that p_i/G_i is constant across products, which is the main analytic result. The particular value that p_i/G_i assumes is simply as algebraic consequence of the zero profit condition.

B. Marginal Cost of Production

It is immediate from (2.4) and (2.8) that the following relationship always holds between quality adjusted marginal cost and the marginal cost of production.

$$G_i(x) = \left[1 + \frac{1}{\eta_i(x)}\right] \frac{\partial F}{\partial x_i}(x, q(x)) \tag{4.10}$$

Recall that $\eta_i(x)$ is the elasticity of demand for product i with respect to intensity expenditures on product i. It will simply be referred to as the intensity elasticity of demand for product i. As explained in section 2, G_i equals the marginal cost of production plus the marginal cost of demand inducement. If the intensity elasticity of product i is high, then increased expenditures on q_i cause large increases in demand. This means that the marginal cost of demand inducement is low, which in turn means that G_i should be closer in value to $\partial F/\partial x_i$. This property is true according to (4.10).

Substitution of (4.10) into (4.7) yields an equivalent statement to (4.7) which is the main result of this part.[14] It will be stated as proposition 4.2.

Proposition 4.2

Suppose that (x,p) is an equilibrium. Then,

$$\frac{p_i / \frac{\partial F}{\partial x_i}(x, q(x))}{1 + \frac{1}{\eta_i(x_i)}} = \frac{p_j / \frac{\partial F}{\partial x_j}(x, q(x))}{1 + \frac{1}{\eta_j(x_j)}} \quad \text{for every } i,j\ . \tag{4.11}$$

That is, the ratio $p_i/(\partial F/\partial x_i)$ is not constant across products. Rather it varies in proportion to $1 + 1/\eta_i$. In particular, if η_i is larger, then $p_i/(\partial F/\partial x_i)$ will be smaller. That is, products with larger intensity elasticities of demand will exhibit lower ratios of price to marginal production cost.

Just as in part A, the fact that profits must be zero in equilibrium actually determines the value of the ratios in (4.11). That is, (4.11) and (4.8) are equivalent to

$$\text{(4.13)} \qquad \frac{p_i / \dfrac{\partial F}{\partial x_i}\big(x, q(x)\big)}{1 + \dfrac{1}{\eta_i(x)}} = \frac{1}{S(x)} \qquad \text{for every } i \ .$$

Therefore corollary 4.2 is equivalent to proposition 4.2.

Corollary 4.2

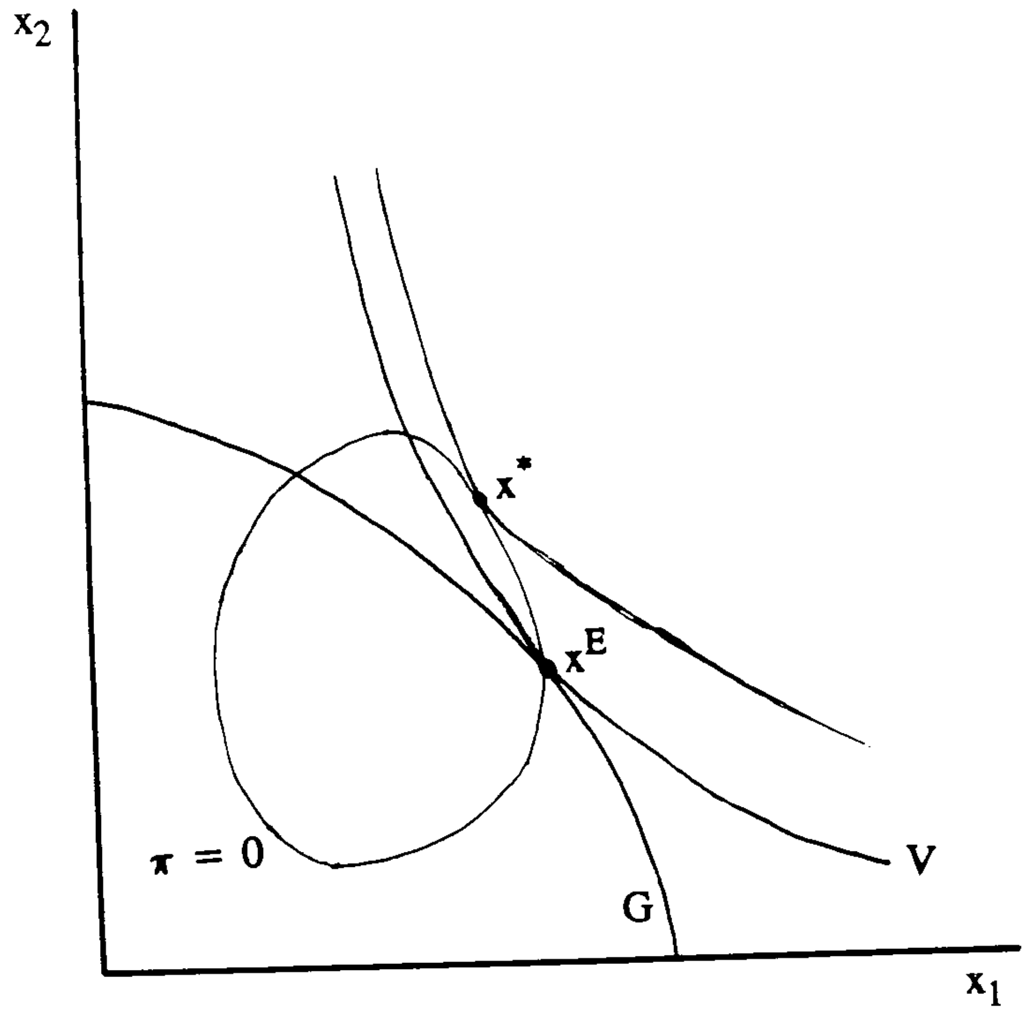
Suppose that (x,p) is an equilibrium. Then (4.13) is true.

Just as in part A, it will be useful to view proposition 4.2 as the statement of the main result. This focusses attention on the result that $p_i/(\partial F/\partial x_i)$ is proportionate to $1 + 1/\eta_i$, which is the main analytic result. The particular value that $[p_i/(\partial F/\partial x_i)]/[1 + 1/\eta_i]$ assumes is then a simple algebraic consequence of the zero profit condition.

As the proof preceding proposition 4.2 makes clear, the explanation for proposition 4.2 is that it is simply an equivalent way of stating the relationship between p and x stated in proposition 4.1 (i.e., given that (4.10) is true by definition, (4.7) and (4.11) are equivalent statements).

Some extra intuition for the result can be had by considering "what goes wrong" if one tries to set prices proportionate to the marginal costs of production. Consider the two-good case described in connection with figure 4.1. Figure 4.2 repeats all of the details shown in figure 4.1 except for the zero profit line which induces x^E. Suppose that $\eta_2 > \eta_1$. Now suppose that the payer chooses prices such that they are proportionate to marginal production cost and such that profits are zero at x^E.

Figure 4.2
Output Choice When Prices are Set Proportional
to Marginal Production Cost At x^E



Thus x^E is feasible for the firm since it will earn zero profits if it chooses x^E. Will the firm choose x^E? The answer is no. The iso-profit line at x^E will be steeper than the iso-cost line.[15] Therefore when the hospital attempts to maximize the utility of health care subject to the zero-profits constraint, it will choose the point x^* which yields a higher level of utility than does x^E.

At first, this result may seem counter-intuitive. The payer chooses prices such that the hospital can break even by choosing x^E. The hospital's goal is to maximize V(x) subject to breaking even. The point x^E maximizes V(x) subject to expenditures being no greater than $G(x^E)$. (In fact it maximizes V(x) - G(x).) Why, therefore, doesn't the hospital choose x^E? The answer is that the hospital can cause the payer to increase expenditures above $G(x^E)$ by choosing x^*. This generates a higher level of benefits from health care. Since the hospital only cares about benefits, it thus prefers x^* Since the payer cares about benefits minus costs, it prefers x^E.

It is useful to describe the above example in terms of the hospital's quality choices. Let $q^* = q(x^*)$ and $q^E = q(x^E)$. Since $x_2^* > x_2^E$, it must be that $q_2^* > q_2^E$. Similarly since $x_1^* < x_1^E$, it must be that $q_1^* < q_1^E$. Now suppose that the principal hopes to induce the agent to choose q^E. Suppose he sets prices proportionate to marginal production cost at $x(q^E)$ and such that the firm's profits would be zero if it chose q^E. The hospital responds by raising intensity on good 2, which has a higher intensity elasticity of demand, and lowering intensity on good 1, which has a lower intensity elasticity of demand. In order to induce the hospital to choose q^E, the payer must choose prices such that the ratio of price to marginal production cost is higher for the good with the lower intensity elasticity of demand.

Thus, one intuitive way to think of proposition 4.2 is that the firm is "biased" towards increasing intensity on products with high intensity elasticities. Thus, in order to induce correct incentives, the payer must choose lower price/marginal cost margins on goods with higher intensity elasticities.

C. The Separable Constant Returns to Scale Case

It is useful to explicitly describe the implications of proposition 4.2 for the separable, linear case of (2.5) since the conclusions are particularly striking in this simple example. Although the formal analysis viewed quantity as the choice variable of the hospital, it will be more natural to view the hospital as choosing intensity in this part.

As explained in part D of section 2, the intensity adjusted cost function exhibits decreasing returns to scale for every x for this case. Therefore, the efficient quantity, x^E, is inducible and will be the equilibrium. Let q^E denote the intensity which causes patients to demand x^E.

Suppose that the payer decides to induce q^E. One possibility to consider would be to set each price equal to the unit cost at q^E, i.e., to choose

$$(4.14) \qquad p_i = c\left(q_i^E\right) \quad .$$

Since $c(q_i^E)$ is the marginal cost of production for good i, prices are thus chosen to be proportional to marginal production cost. The hospital can break even by choosing q^E. The hospital is completely altruistic and q^E is the efficient intensity. Why, then, doesn't the hospital choose q^E? The answer is that, so long as intensity elasticities of demand vary across products, the hospital can raise intensity on goods with higher elasticities and lower intensity on goods with lower elasticities and simultaneously

(i) still break even,

and (ii) increase the payer's expenditures.

This increase in expenditures increases the benefits of health care so the hospital prefers this outcome. Of course, costs increase by more than benefits, so the payer prefers q^E.

This does not mean that q^E is unattainable. The payer can induce the hospital to choose q^E by setting prices according to (4.13). That is, prices should be set so that $p_i/c_i(q_i^E)$ is proportional to $(1 + 1/\eta_i(x_i^E))$ and the firm can just break even at q^E. Formally, p should be chosen to satisfy[16]

$$\text{(4.15)} \qquad \frac{p_i}{c_i\left(q_i^E\right)} = k\left(1 + \frac{1}{\eta_i\left(x_i^E\right)}\right) \text{ for some k and}$$

$$\text{(4.16)} \qquad \sum_{i=1}^{n} \left[p_i - c_i\left(q_i^E\right)\right] x_i^E = 0 \quad .$$

Note that, if $\eta_i(x_i^E)$ is not constant for every i, then (4.15) implies that $p_i/c_i(q_i^E)$ will not be constant for every i. Then (4.16) implies that $p_i/c_i(q_i^E)$ will be greater than 1 for some values of i and less than 1 for others. Thus, both negative and positive profit margins will be observed in equilibrium. Thus, far from being the signal that something is remiss, the existence of positive and negative profit margins is the expected result when all is well and the payer is managing to induce the hospital to choose the efficient level of care, q^E.

D. Ramsey Pricing

This paper's "inverse elasticity" rule resembles the familiar Ramsey pricing rule from regulation theory.[17] Thus it is of some interest to explore the connection between the two results. Although the results are closely connected, the nature of the connection is not the one suggested by the superficial resemblance of the inverse elasticity rules. In particular, there does not appear to be a simple relabelling of the variables in the Ramsey problem (for example, by replacing price in the standard Ramsey problem with intensity) that directly yields the inverse elasticity result of this paper. Rather, the connection is somewhat more indirect.

The main result of this paper is proposition 4.2, which states that the ratio of price to marginal production cost varies inversely with intensity elasticity across products. This result follows directly from proposition 4.1 and equation 4.10. Proposition 4.1 states that the ratio of price to intensity adjusted marginal cost is constant across products. Equation 4.10 states the relationship between the intensity adjusted marginal cost and the production (or constant intensity) marginal cost.

It will now be shown that the result of proposition 4.1 can be interpreted as following from a version of the Ramsey problem.

The general Ramsey problem is

(4.17) Max $W(x)$

(4.18) subject to $\sum_{i=1}^{n} p_i(x)x_i - C(x) \geq 0$

where W is the social utility function, $p_i(x)$ is the inverse demand curve for good i, and C is the cost function. The general result, is that the optimal choices must satisfy

(4.19) $$W_i(x) + \lambda \left[\sum_{j=1}^{n} \frac{\partial p_j}{\partial x_i} x_j + p_i - C_i \right] = 0 \ .$$

Problem (4.17)-(4.18) is the hospital's problem where W(x) equals V(x), C(x) equals G(x), and $p_i(x)$ is taken as the fixed price, p_i. Substitution of these values into (4.19) yields

(4.20) $$V_i(x) + \lambda[p_i - G_i] = 0 \ .$$

Note that all of the derivatives of price with respect to output vanish, because, in the hospital's problem, price is fixed. Now, when the payer performs his optimization, he induces an x such that

(4.21) $$V_i = G_i$$

Substituion of (4.21) into (4.20) yields proposition 4.1.

In summary, the intermediate conclusion, proposition 4.1, that prices are proportional to intensity adjusted marginal cost can be viewed as being derived from the Ramsey problem. There does not appear to be a simple relabelling of the variables in the Ramsey problem that directly yields the final conclusion, proposition 4.2.

5
Accounting Cost

Proposition 4.2 characterizes the nature of the ratio of price to marginal production cost across products. Unfortunately, the only readily available cost data for most hospitals is accounting cost. There is, in fact, a large literature in health care policy which is concerned with comparing the ratio of price to accounting cost across products. The focus is on identifying whether cross-subsidies occur when accounting cost is viewed as the "true" cost of each product.[18] This literature often refers to itself as studying cost-shifting.

Therefore, it would be extremely useful if the results of this paper could be interpreted as applying to the ratio of price to accounting cost. In particular, it would be useful if proposition 4.2 were still true when the marginal cost of production is replaced by the accounting cost of production. In this case, data would be readily available to test or apply the theory. Furthermore, the theory could be interpreted as providing a new analytic perspective on the cost-shifting literature. The major result of this section will be to show that relatively natural assumptions exist which generate this result.

Define an accounting system to be a function from x to R^n denoted by $a(x) = (a_1(x),\ldots, a_n(x))$ where $a_i(x)$ is interpreted as the unit accounting cost of product i given that the hospital produces x (and, as usual, given that the hospital chooses q so that patients demand x). That is, in order to produce the quantity x at treatment intensities q(x), the hospital incurs various expenditures. The hospital's accounting system assigns expenditures to products. The total expenditures assigned to product i divided by the units produced of product i, equals $a_i(x)$, the unit accounting cost of product i.

Formally, then, the goal of this section is to identify a sufficient condition for the statement

(5.1) (x,p) is an equilibrium $\Rightarrow$ $$\frac{p_i/a_i(x)}{1+\dfrac{1}{\eta_i(x)}} = \frac{p_j/a_j(x)}{1+\dfrac{1}{\eta_j(x)}} \quad \text{for every i,j}$$

to be true. This means that a_i can be substituted for $\partial F/\partial x_i$ and proposition 4.2 is still true.

By comparing (4.11) and (5.1), it is clear that a sufficient condition for this is that accounting costs be proportional to marginal costs, i.e., that

(5.2) $$a_i(x) = k\frac{\partial F}{\partial x_i}(x,q(x))$$

for every i for some

(5.3) $$k > 0 \quad .$$

This result is stated as proposition 5.1.

Proposition 5.1

Suppose that a(x) satisfies (5.2)-(5.3). Then if (x,p) is an equilibrium, (5.1) is true.

A relatively plausible example exists which satisfies the sufficient conditions (5.2)-(5.3). Suppose that variable costs are linear and separable as described by (2.5) but that there may be a fixed cost as well. Formally, assume that costs are given by

(5.4) $$F(x,q) = K + \sum_{i=1}^{n} c_i(q_i)x_i$$

where $K \geq 0$ and the second term on the RHS of (5.4) satisfies (b.1) and (b.2). Assume that the firm calculates the accounting cost of product i by directly assigning the variable cost and then allocating the fixed cost across products in proportion to variable costs. Then the total accounting cost of product i equals

$$(5.5) \qquad c_i(q_i)x_i + \frac{Kc_i(q_i)x_i}{\sum_{j=1}^{n} c_j(q_j)x_j} .$$

The unit accounting cost of product i is obtained by dividing (5.5) by x_i and equals

$$(5.6) \qquad a_i(x) = c_i(q_i(x_i)) \left\{ 1 + \frac{K}{\sum_{j=1}^{n} c_j(q_j(x_j))x_j} \right\} .$$

The marginal cost of production equals $c_i(q_i(x_i))$ and the term in brackets does not depend on i. Therefore, this example satisfies (5.2)-(5.3).

Note that if K equals zero, then $a_i(x)$ is then set precisely equal to $c_i(q_i(x_i))$. Since accounting cost is set precisely equal to marginal cost, obviously proposition 4.2 holds true for accounting costs for this case. The above example shows that proposition 4.2 continues to hold true for accounting costs when there are fixed costs, so long as the fixed costs are allocated in proportion to the variable costs.

To summarize the above, suppose that the following three conditions are satisfied.

(i) Variable costs are linear and separable.

(ii) The hospital correctly assigns variable costs to products.

(iii) Fixed costs are allocated in proportion to variable costs.

Then (5.1) is true. That is, accounting costs can be substituted for marginal costs and proposition 4.2 is still true.

Of course, none of the above three conditions is likely to be perfectly and completely satisfied in real situations. However, all three conditions are relatively plausible and may well be "close to true" in many situations. Therefore, it may be that the desired result is approximately correct in many situations. This is basically an empirical issue and therefore cannot be further resolved by this paper.

6
Expenditure Caps

This paper's model suggests a possible role for expenditure caps in a prospective payment system. Define an expenditure cap to be a non-negative real number, M. Now assume that the payer chooses a pair (p,M). This means that the hospital will receive revenues according to price vector, p, but only up to a maximum of M. Formally, let R(x,p,M) denote the hospital's revenues if patients receive x from the hospital. It is defined by

$$(6.1) \qquad R(x,p,M) = \min\left\{\sum_{i=1}^{n} p_i x_i, M\right\}$$

It will be said that (p,M) induces x if the hospital chooses x given (p,M). Formally (p,M) induces x if x solves

$$(6.2) \qquad \underset{x>0}{\text{Maximize}} \quad V(x)$$

$$(6.3) \qquad \text{subject to} \quad R(x,p,M) - G(x) = 0 \ .$$

The value of an expenditure cap now can be explained. Recall that x^E is the efficient outcome. In the model of the paper with no expenditure caps, x^E was inducible if and only if $S(x^E) \geq 1$. When x^E was inducible, there was a unique price that induced it. Suppose that an expenditure cap is set equal to the cost of the efficient outcome, i.e.,

$$(6.4) \qquad M = G(x^E) \ .$$

Suppose also that p is chosen to be any price vector such that the firm breaks even at x^E, i.e.,

$$(6.5) \qquad \sum_{i=1}^{n} p_i x_i^E - G(x^E) = 0 \ .$$

Then (p,M) induces x. This result is formally stated as proposition 6.1.

Proposition 6.1

Suppose that (p,M) satisfies (6.4) and (6.5). Then (p,M) induces x^E.

Before explaining the significance of this result, its intuition will be outlined. In the model with no expenditures, the reason that $p^*(x^E)$ was the unique price that induced x^E was that, for any other price vector such that the hospital broke even at x^E, it could choose a different x and induce the payer to increase his expenditures. This is no longer possible with an expenditure cap.

The result in proposition 6.1 has two important features. By far the most important feature is that any price which causes the firm to break even at x^E can be used to induce x^E. This means that prices could be chosen to satisfy some other criterion. For example, prices could be chosen so that price equals accounting cost for each product if this was thought to be desirable. In a more general model where price affects patient demand and thus directly influences welfare, prices could be chosen to help solve this welfare problem. This feature also means that less information is required on the payer's part to induce x^E. Without expenditure caps, the payer had to calculate the particular price vector $p^*(x^E)$ and this depended upon the intensity elasticities. When the payer can choose any price vector such that the hospital breaks even, he no longer has to know the value of intensity elasticities.

The second feature of the result is that x^E is inducible even if $S(x) < 1$. Thus output vectors where G(x) exhibits increasing returns to scale can be induced if this is efficient.

7
Alternate Objective Functions

The purpose of this section is to consider three alternate assumptions about the nature of the hospital's objective function and show the basic result that the ratio price to marginal production cost will vary inversely with intensity elasticity continues to hold.

A. Different Preferences Over Health Care

First, suppose that, while the hospital has preferences over health care, that these preferences are not necessarily identical to the payer's. Formally, assume that V^P is the payer's utility function and V^H is the hospital's utility function, and both functions satisfy the assumptions made for V. Then lemma 4.1 is still true for V^H. Substitution of (4.10) into (4.11) yields the following result.

Proposition 7.1

Suppose that the hospital's preferences are described as in part A of this section. Suppose that p induces x. Then

$$p_i \,/\, \frac{\partial F}{\partial x_i}\big(x, q(x)\big) = \left[1 + \frac{1}{\eta_i(x)}\right] \left[1 - \frac{V_i^H(x)}{G_i(x)\delta^*(x)}\right] \quad . \tag{7.1}$$

Unfortunately, it is no longer necessarily the case that the ratio $V_i^H(x)/G_i(x)$ will equal 1 when $x = x^E$. This is because x^E maximizes $V^P(x) - G(x)$. This means that, even in equilibrium, it may be that $p_i/(\partial F/\partial x_i)$ varies because of variations in V_i^H/G_i.

Therefore, the best that can be said in this case is that $p_i/(\partial F/\partial x_i)$ will be affected by two factors. It will vary inversely with the intensity elasticity of demand, η_i, and inversely with the hospital's own assessment of the marginal utility per dollar from increasing x_i. Thus, although $p_i/(\partial F/\partial x_i)$ still varies inversely with η_i, this is only one of two factors that will influence $\partial F/\partial x_i$.

B. Revenue Maximization

Another possibility is that the nonprofit hospital is not as altruistic as described in the main model. Instead of having preferences over health care and attempting to maximize the benefits of health care, the hospital may simply want to grow as large as possible. The assumption that bureaucrats wish to maximize the size of their organization for various reasons (prestige, career enhancement, etc.), is a typical one in the economics literature. A natural measure of size in this model is the hospital's revenues. Formally then, p induces x if x solves

$$(7.2) \qquad \underset{x \geq 0}{\text{Maximize}} \quad \sum_{i=1}^{n} p_i x_i$$

$$(7.3) \qquad \text{subject to} \quad \sum_{i=1}^{n} p_i x_i - G(x) = 0$$

It follows immediately from the first order conditions that the desired relation must hold. This result is stated as proposition 7.2.

Proposition 7.2

Suppose that x solves (7.2)-(7.3) for p and that $x > 0$. Then (4.11) is true.

Note that this is a stronger result than proposition 4.2. Proposition 4.2 required (x,p) to be an equilibrium in order to guarantee that (4.11) is true. Proposition 7.2 only requires that p induce x.

The intuition for proposition 7.2 is as follows. As explained in section 4, in order to induce some x, price must be chosen so that the hospital cannot increase revenues and still break even. This is program (7.2)-(7.3).

C. Profit and Revenue Maximization

This part will assume that the hospital is even less altruistic than in part B. It will be assumed that the hospital values both revenues and profit and attempts to maximize a utility function which depends on both. The motivation for this assumption is the idea that managers of a nonprofit may still value profit if it can be converted to managerial perks, higher salaries, or used to fund projects of particular interest to managers. The assumption that a nonprofits goal is to maximize a utility function depending upon profit and other variables is, in fact, quite standard in the literature (Hodgkin and McGuire 1993, Wiesbrod 1993).

Formally, then the hospital chooses x to solve

$$(7.4) \qquad \text{maximize} \quad \Gamma(\pi, R)$$

where

$$(7.5) \qquad \pi = \sum_{i=1}^{n} p_i x_i - G(x)$$

$$(7.6) \qquad R = \sum_{i=1}^{n} p_i x_i$$

and Γ is a twice continuously differentiable function strictly increasing in π and weakly increasing in R.

Just as in part B, the first order conditions immediately imply the desired result which is stated as proposition 7.3.

Proposition 7.3

Suppose that x solves (7.4) and $x > 0$. Then (4.11) is true.

Just as for proposition 7.2, this is a stronger result than proposition 4.2 because proposition 7.3 only requires that p induce x in order for (4.11) to be true.

Also note that proposition 7.3 applies to the case of a pure profit maximizing firm. This is the special case where $\Gamma(\pi,R)$ equals π. The model of a profit-maximizing monopolist or oligopolist choosing product quality is well studied in the economics literature.[19] Equation (4.11) is obtained by manipulating the standard first order condition to this problem. However, there does not appear to be a paper that makes the precise statement that (4.11) is true when x satisfies (7.4).

8
Discussion

Part A will discuss the implications of this paper for interpreting the behavior of price/cost ratios in hospitals. Then part B will explain how this paper's theory provides a different explanation for differential price/cost ratios across payer groups than the traditional "cost shifting" explanation. Finally, part C will discuss empirical predictions and tests.

A. Price/Cost Ratios

For the purposes of this discussion, it will be assumed that accounting costs are relatively proportional to marginal costs for most hospitals so that the conclusions of this paper apply to price/cost ratios when cost is measured by unit accounting cost. Therefore the term "price/cost ratio" should be interpreted as the ratio of price to accounting cost for the remainder of part A.

The major conclusion of this paper is that price/cost ratios are endogenously determined by a hospital's treatment intensity choices even if prices are completely fixed. Price/cost ratios will vary across products inversely with the products' intensity elasticity of demand. That is, price/cost ratios will be higher on products where demand is less responsive to increased intensity. Therefore, in an accounting sense, the hospital will be earning positive profits on some products (those with low intensity elasticities of demand) and negative profits on some products (those with high intensity elasticities of demand).

From a policy perspective, the major contribution of this paper may be that it suggests a very different theory of how a nonprofit hospital chooses treatment intensities than the traditional theories. This, in turn, implies a different normative theory of how a payer should choose prices. There are perhaps two traditional theories of how a nonprofit hospital chooses treatment intensities. The first, which will be called the "exogenous intensity" theory is that hospitals determine treatment intensity based on medical issues completely independently from prices. The second theory, which will be called the "endogenous but separable" theory is that, for each product, the hospital increases treatment intensity up to the point that price just equals the accounting cost of supplying the product. Thus the hospital breaks even on each product.

Under the traditional view, it has generally been thought to be desirable that all price/cost ratios be equal to one to avoid cross-subsidies. This is thought to be desirable based on fairness grounds.

Both of the above theories, together with the view that cross-subsidies are undesirable, generate relatively similar normative theories of price setting. Under the exogenous intensity theory, the payer begins by simply estimating the exogenous intensity choices of the hospital. Then, given the view that cross-subsidies are undesirable, the unambiguous best choice is to set the price of each product equal to its cost given the treatment intensities. Under the slightly more sophisticated endogenous but separable theory, the payer begins by choosing the vector of treatment intensities that he wishes to induce. He then induces the desired vector by setting each price equal to accounting cost given the desired treatment intensity. This policy has the desirable side effect that there are no cross-subsidies.

This paper shows that neither of these theories of intensity choice stand up very well to logical scrutiny. The theory that intensity choice is completely independent of all financial issues is, of course, flawed. If a hospital must break even to survive and is legally prevented from earning positive profits, then intensity choice must respond to financial issues. The "endogenous but separable" theory is slightly more sophisticated in that it acknowledges this point. The theory that a hospital will adjust intensity to just break even on each product might be superficially appealing. If a hospital is a nonprofit it must break even. If the hospital is altruistic, in the sense that it agrees with the payer on the relative social value of increasing intensity across products, it may seem appealing to conclude that the hospital will "obey the payer's command" and choose intensity on each product to break even.

This paper shows that, while perhaps superficially appealing, this logic is incorrect. The logically correct statement of an altruistic nonprofit hospital's decision problem is that it attempts to maximize the social value of health care (using the payer's definition of social value) subject to the constraint that it break even. It is shown that the solution to this problem generally does not involve the hospital breaking even on each product. Relative to the intensity choice that would cause the hospital to break even on a product by product basis, the hospital increases intensity on products with high intensity elasticities and decreases intensity on products with low intensity elasticities. The reason for this is that the hospital can increase the payer's total expenditures on health care by doing this and thus increase the social utility of health care. This paper's theory suggests the following normative theory for price setting by the payer. First, the payer should decide on a desirable level of intensity for each product and estimate its accounting cost.[20] Then the payer should estimate the intensity elasticity of demand, η_i, for each product. Finally, the payer should set prices such that the price/cost ratios are proportional to $1 + 1/\eta_i$ and the firm will just break even when it chooses the desired intensity vector.

Note that in this paper's model, prices can no longer perform the role of eliminating cross-subsidies. So long as intensity elasticities vary across products, then so must price/cost ratios. In the formal model of this paper, this is of no great concern because there is no reason to value the elimination of cross-subsidies. The concept of a "cross-subsidy" is simply an accounting artifact and has no direct bearing on anyone's welfare. This is because there is a single payer and individual patients are fully insured. Furthermore, even in more complex models with multiple payers or patients that partially pay for services, it seems likely that appropriate lump-sum payments could be made between payers or from the payer to the patient to achieve any desired surplus division among payers or among patients.

Therefore it seems plausible that, in many circumstances, the issue of whether cross-subsidies exist is of no great welfare consequence. The view that prices should equal costs on a product-by-product basis may stem largely from the fact that there has been no good theory to explain how prices should be set and, in the absence of any good theory, the simple idea that price should equal cost seemed appealing. This paper provides a theory to fill this vacuum.

It may be in more complex models that the choice of price could be modelled as having some direct welfare consequences. For example, if patients pay for a fraction of their health care themselves, then price will influence patients' demand and this will have direct welfare consequences. In such a model these extra welfare effects would have to be taken into account, and the solution of this paper would no longer necessarily be optimal. Nonetheless, the factors identified in this paper should continue to play a role in more complex situations. The discussion of expenditure caps in section 6 is also very relevant to this issue. In particular, imposition of expenditure caps creates a "degree of freedom" to choose relative prices to solve some other problem.

B. Cross-Subsidies Across Payer Groups and Cost Shifting

The most remarked-upon instance of cross-subsidies is between Medicare and privately insured patients groups. On average, the prices hospitals receive for supplying services to Medicare patients are less than the accounting costs of the services provided, while the prices hospitals receive for supplying services to privately insured patients are greater than the accounting costs of the services provided.[21] The accepted explanation for this phenomenon is that hospitals have responded to low Medicare prices by raising the prices they charge to privately insured patients. This behavior is usually referred to as "cost shifting." One difficulty with this explanation is that it does not explain why large private insurers have accepted these price increases.

The purpose of this part is to argue that this paper's model can be interpreted to supply an alternate explanation of this phenomenon which does not rely on the lack of bargaining power of large private insurers. Just as in part A, assume that accounting costs are proportional to marginal costs so that this paper's theory applies to accounting costs.

Assume that the hospital's objective is as modelled in part C of section 7. That is, its goal is to maximize a utility function depending upon both profit and expenditures. Therefore, by proposition 7.3, the basic inverse elasticity result, equation (4.11), holds for any fixed price vector, so long as the hospital maximizes its utility given the fixed price vector.[22]

Assume that there are two payers, the federal government and a private insurer. For simplicity, assume that there are two groups of patients and each group consumes a single health care product. Let product 1 be health care for privately insured patients and let product 2 be health care for Medicare patients. Assume that an individual patient's demand for health care as a function of the treatment intensity offered is identical across all patients and that the intensity elasticity of demand decreases as intensity grows larger. Assume that a (nonmodelled) strategic interaction between the two payers determines an equilibrium price vector (p_1^*,p_2^*). Then the hospital chooses the intensity vector (q_1^*,q_2^*) to maximize its utility given the price vector (p_1,p_2). Consistent with observed outcomes, assume that, in the resulting equilibrium, privately insured patients receive a higher intensity of treatment than do Medicare patients, i.e., $q_1^* > q_2^*$. This outcome presumably occurs in equilibrium because the payer for the privately insured patients is willing to pay more for intensity increases than is the federal government.

Since all patients are alike and intensity elasticity declines with intensity, this implies that the demand curve of the privately insured group will exhibit a lower intensity elasticity of demand than the demand curve of the Medicare insured group at the induced intensity choices. The major result of this paper (proposition 7.3, for this version of the model) is, of course, that price/cost ratios vary inversely with intensity elasticity. I particular, therefore, it must be true that there is a higher price/cost margin on the privately insured group than the Medicare insured group. If the hospital is earning zero profit overall, it must therefore be earning a positive accounting profit on the privately insured group and a negative accounting profit on the Medicare insured group.

Therefore, in the above model, the hospital earns a positive accounting profit on privately insured patients and a negative accounting profit on Medicare patients even though the hospital has absolutely no ability to affect prices. The profit differentials do not result from the hospital raising prices to private insurers (as in the standard "cost-shifting" explanation). Rather, they result from the hospital's choice of treatment intensities in response to fixed prices.

Note that this paper's explanation for observed profit differentials depends on the property that the intensity elasticity of demand is higher for Medicare patients than privately insured patients. This simple property is sufficient to generate the observed differentials through its effects on the hospital's intensity choices. I am not aware of any empirical investigations of intensity elasticities across patient groups. This paper clearly suggests that such empirical measurements would be interesting.

C. Empirical Predictions and Tests

Although the variation of price/cost ratios across payer groups is an appealing issue to apply this paper's theory to (because of the great policy interest in this issue), the most natural issue to apply this paper's theory to is price/cost margins across products for a single payer. In particular, the case of variation in price/cost margins across DRGs for Medicare is a very natural candidate to apply this paper's model to. This is because there is definitely a single payer and prices are definitely fixed by the payer. This case, in fact, supplies an opportunity to empirically test predictions of this paper's theory.

Two different possible approaches to testing this paper's theory are possible. The more ambitious approach would be to attempt to directly estimate intensity elasticities and then directly test either (4.11) or (5.1) depending upon whether estimates of marginal cost or accounting cost were available.

The second, less ambitious but much more feasible approach would be to identify observable characteristics of products that would tend to indicate that intensity elasticity is relatively high or low.

One would then test the proposition that price/cost ratios tend to be higher for products that exhibit characteristics which suggest that their intensity elasticities are lower.

One obvious example of such a characteristic is the presence or absence of competition for a particular product. For example, suppose that two hospitals, A and B, are situated relatively close to one another. Suppose that both hospitals supply product 1 but only hospital A supplies product 2 and only hospital B supplies product 3. Then it seems plausible that product 1 may exhibit a much higher intensity elasticity than products 2 or 3 (i.e., a slight increase in product 1's intensity may cause a large demand increase because all patients will shift to the hospital supplying higher intensity). Thus, this paper's model suggests that price/cost ratios will be lower on product 1 than products 2 or 3. A similar type of analysis could be conducted for a single DRG across different hospitals. It may be that some hospitals experience more competition in a particular DRG than do other hospitals. This paper predicts that price/cost ratios would be lower for hospitals experiencing more competition.

Notes

1. See Prospective Payment Assessment Commission (1993) for a more detailed history. See Glazer and McGuire (1993a) for statistics on the use of the prospective payment method.

2. See Wiesbrod (1993) for a general discussion of the possible goals of nonprofits, some empirical evidence on this issue, and references to the literature.

3. See Dor and Farley (1991), Glazer and McGuire (1993), Ma and Burgess (1993), Ma and McGuire (1993), Spence (1975), and Wolinsky (1993).

4. See Ma and McGuire (1993) for an interesting analysis of strategic interaction between payers

5. A discharge is the treatment of one patient for one diagnosis.

6. The analysis of this paper remains similar if demand is allowed to depend on price as well as treatment intensity. Since the hospital views prices as fixed when choosing intensity, the conditions characterizing the hospital's choice of intensity remain unchanged. However, the payer's choice of prices is now complicated by the additional factor that prices affect the level of demand as well as the hospital's intensity choice. This additional factor complicates some of the conclusions.

7. Whether a hospital might find it optimal to purposely engage in rationing is a topic beyond the scope of this paper.

8. Throughout this paper for any vector $y = (y_1, \ldots, y_n) \in R^n$, the notation $y > 0$ will be used to denote the property that $y_i > 0$ for every i and the notation $y \geq 0$ will be used to denote the property that $y_i \geq 0$ for every i.

9. For functions of x, the subscript i will be used to denote the derivative with respect to x_i.

10. Since these assumptions are only made for the special case and are not maintained throughout the paper they will be numbered differently than the other assumptions. When the special case (2.5) is referred to in the remainder of the paper, it will always be assumed that (b.1) and (b.2) hold as well.

11. See Baumal, Panzar, and Willig (1982), p. 50.

12. See Friedman and Pauly (1981).

13. All proofs are contained in an appendix.

14. Obviously versions of lemmas 4.1-4.3 can also be derived by substituting (4.10) into the various results.

15. It is straightforward to algebraically show this is true because $\eta_2 > \eta_1$.

16. In this special case, η_i depends only on x_i and not the entire vector x as shown in (2.8). Therefore, η_i is written only as a function of x_i.

17. See Spulber (1989) for a treatment of Ramsey pricing and references to the literature.

18. See Carey and Stefos (1992), Congressional Budget Office (1993) and Prospective Payment Assessment Commission (1993) for a discussion and further references.

19. See, for example, Dor and Farley (1991), Ma and Burgess (1993), Spence (1975), and Wolinsky (1993).

20. Of course this paper's theory also explains how the payer determines the desirable level of intensity.

21. See Congressional Budget Office (1993) and Prospective Payment Assessment Commission (1993).

22. The advantage of this formulation over the formulation used in the basic model, is that, in the basic model, the desired result only holds when the payer chooses price optimally. In the interpretation of this part there are two payers and there is no reason to believe that they choose a jointly optimal price vector.

References

Baumol, William, John Panzar, and Robert Willig (1982), "Contestable Markets and the Theory of Industry Structure," Harcourt Brace Jovanovich, Inc., 1982.

Carey, Kathleen and Theodore Stefos (1992), "Managing Inpatient and Outpatient Costs: A Cost Function Approach," *Health Care Financing Review*, 14, 115-124

Dor, Avi and Dean Farley (1991), "The Cost and Quality of Health Care by Source of Payment: Inferences from a Multi-Product Cost Function," mimeo, Agency for Health Care Policy and Research, Rockville, MD.

Friedman and Pauly (1981), "Cost Functions for a Service Firm With Variable Quality and Stochastic Demand: The Case of a Hospital," *Review of Economics and Statistics*, 63(4), 620-624.

Glazer, Jacob and Thomas McGuire (1993a), "Payer Competition and Hospital Payment Schemes," mimeo, Department of Economics, Boston University.

Glazer, Jacob and McGuire, Thomas (1993b), "Should Physicians be Permitted to 'Balance Bill' Patients?" *Journal of H ealth Economics*, forthcoming.

Harris, Jeffrey (1979), "Pricing Rules for Hospitals," *Bell Journal of Economics*, 10, 224-243.

Hodgkin, Dominic and Thomas McGuire (1993), "Payment Levels and Hospital Response to Prospective Reimbursement," *Journal of Health Economics*, forthcoming.

Ma, Albert and James Burgess, Jr. (1993), "Quality Competition, Welfare and Competition," *Zeitschrift for Nationalokonomie Journal of Economics*, forthcoming.

Ma, Albert and Thomas McGuire (1993), "Paying for Joint Costs in Health Care," *Journal of Economics and Management Strategy*, forthcoming.

Prospective Payment Assessment Commission (1993), *Medicare and the American Health Care System: Report to the Congress.*

Rogerson, William (1990), "Quality vs. Quantity in Military Procurement," *American Economic Review*, 80, 83-92.

Rogerson, William (1991), "Incentives, the Budgetary Process, and Inefficiently Low Production Rates in Defense Procurement," *Defence Economics*, 3, 1-18.

Spence, Michael (1975), "Monopoly, Quality, and Regulation," *Bell Journal of Economics*, 6, 417-29.

Spulber, Daniel (1989), *Regulation and Markets*, MIT Press, Cambridge.

Weisbrod, Burton (1993), "Nonprofit and Propietary Sector Behavior: Does Institutional Form Matter," mimeo, Department of Economics, Northwestern University.

Wolinsky, Asher (1993), "Regulation of Duopoly Under Asymmetric Information," mimeo.

Appendix

Lemma 3.1

First suppose that x satisfies (3.1)-(3.2). It is obviously sufficient to show that (3.2) is satisfied with equality. Suppose, for contradiction, that profit was strictly positive. Then by continuity it would also be strictly positive for a small increase in output which would increase the objective function. Thus x cannot be a solution.

Now suppose that x satisfies (2.9)-(2.10). Suppose for contradiction that it does not satisfy (3.1)-(3.2). Thus there exists some $\hat{x}$ such that

$$V(\hat{x}) > V(x) \tag{A.1}$$

and

$$\sum_{i=1}^{n} p_i \hat{x}_i - G(\hat{x}) > 0 \quad . \tag{A.2}$$

Assumption (a.6) implies that there must exist a $\theta > 1$ such that

$$\sum_{i=1}^{n} p_i \theta_i \hat{x}_i - G(\theta \hat{x}) = 0 \quad . \tag{A.3}$$

Thus $\theta\hat{x}$ satisfies the zero profit constraint. As well,

$$V(\theta \hat{x}) > V(\hat{x}) \tag{A.4}$$

since $\theta > 1$. Equations (A.1) and (A.4) imply that

$$V(\theta \hat{x}) > V(x)$$

contradicting the assumption that x solves (2.8)-(2.9). QED

Lemma 3.2

Conditions (3.3)-(3.5) are obviously necessary. The question is whether they are sufficient. For $\delta < \infty$, (3.4)-(3.5) are the Kuhn-Tucker conditions. Since G is convex for $x > 0$ the constraint set is convex for $x > 0$. By assumption, the objective function is concave. Therefore the necessary conditions are sufficient.

For $\delta = \infty$, (3.3) is the first order conditions for x to maximize profit. Since G(x) is strictly convex for $x > 0$, x must be the unique maximum to G(x) over $x > 0$. By (3.4), profits at x are zero. Therefore, there is a single $x > 0$ satisfying the break-even constraint. This, therefore, is the solution.

QED

Lemma 3.3

Simply invert (3.3) and (3.4).

QED

Proposition 3.1

The proof is contained in the text preceding the proposition.

QED

Lemma 3.4

For the purpose of this proof define V(x) to be equal to $-\infty$ for any x such that $x_i = 0$ for some i. By (a.3), V can be viewed as a continuous function over $[0,\infty)^n$. Choose any $\hat{x} > 0$. Let k denote

(A.5) $$k = V(\hat{x}) - G(\hat{x}) \quad .$$

By (a.7), the set

(A.6) $$X_k = \{x : x \geq 0 \text{ and } V(x) - G(x) \geq k\}$$

is bounded. It is also closed since V and G are continuous. Therefore, V(x)-G(x) achieves a maximum over this set. This is obviously the unconstrained maximum as well. Since V is strictly concave and G is strictly convex, the maximum is unique.

QED

Lemma 3.5

Define (A.5) and (A.6) just as in lemma 3.4. When doing so, choose an x such that $S(x) \geq 1$. Recall that X_k is compact. The set

$$S = \{x : S(x) \geq 1 \text{ and } x \geq 0\} \tag{A.7}$$

is closed. Therefore $S \cap X_k$ is compact. Therefore V(x) - G(x) achieves a maximum over this set. This maximum is obviously also the maximum to (3.12) - (3.13).

QED

Lemma 3.6

The proof is contained in the text preceding the proposition.

QED

Lemma 3.7

For contradiction, assume that $(\hat{x},p)$ is an equilibrium and that $S(\hat{x}) > 1$. Let x^α denote the value of x which is a convex combination of x and x^E.

$$x^\alpha = \alpha\hat{x} + (1-\alpha)x^E \quad . \tag{A.8}$$

Let f(x) denote the payer's objective function.

$$f(x) = V(x) - G(x) \tag{A.9}$$

By assumption, x^E is the unique maximizer of f. Therefore

(A.10) $$f(x^E) > f(\hat{x}) \quad .$$

By assumptions (a.2) and (a.4), f is strictly concave. Therefore

(A.11) $$f(x^\alpha) > f(\hat{x})$$

for every $\alpha \in (0,1)$ Since S(x) is continuous and $S(\hat{x}) > 1$, $S(x^\alpha) > 1$ for values of α sufficiently close to 1. This contradicts the assumption that $\hat{x}$ solves (3.12)-(3.13).

QED

Lemmas 4.1-4.3, Propositions 4.1-4.2, Corollaries 4.1-4.2

All proofs in section 4 are contained in the text prior to the formal statements of the results.

QED

Proposition 6.1

Program (6.2)-(6.3) can be rewritten as

(A.12) $$\underset{x>0}{\text{Maximize}} \quad V(x)$$

(A.13) $$\text{subject to} \quad \sum_{i=1}^{n} p_i x_i - G(x) = 0$$

(A.14) $$G(x) \leq M$$

Now suppose that (6.4) is true. By definition, x^E solves the program of maximizing (A.12) subject to (A.14). However, by (6.5), x^E satisfies (A.13). Therefore it solves (A.12)-(A.14).

QED

Propositions 7.1-7.3

The proofs of these propositions are contained in the text prior to the statements of the propositions.

QED