

Market Equilibria with Not-For-Profit Firms

Phelps, Charles E. and Itai Sened

Working Paper No. 191
June 1989

University of
Rochester

Market Equilibria with Not-For-Profit Firms

Charles E. Phelps and Itai Sened

Rochester Center for Economic Research
Working Paper No. 191

MARKET EQUILIBRIA WITH NOT-FOR-PROFIT FIRMS

Charles E. Phelps
and
Itai Sened

Department of Political Science
University of Rochester
Rochester, NY 14627

Working Paper No. 191

June 1989

June 28, 1989
First Draft.

comments welcome

MARKET EQUILIBRIA WITH NOT-FOR-PROFIT FIRMS

Charles E. Phelps and Itai Sened

Department of Political Science
University of Rochester
Rochester, NY 14627

ABSTRACT

This paper provides a comprehensive framework to analyze the supply of medical care in a market structure common in the United States. The framework has three novel features: First, it portrays equilibrium quantity, quality, and price in not-for-profit hospital markets, both under conditions of free entry and with regulatory constraints on entry. Second, it shows that many phenomena, commonly attributed to demand inducement and/or explained by "target income" models of physicians' behavior, emerge directly from a standard monopolistic competition model, constrained only by the assumption that consumer search is limited. Third, we link explicitly hospital and physician behavior, showing how each affects the other's cost functions and demand curves. Throughout the paper we provide some stylized evidence to demonstrate the usefulness of our model.

I. Introduction.

The behavior of not-for-profit (NFP) organizations, particularly hospitals, has attracted the attention of economists for decades, with works initially by Arrow (1963), Newhouse (1970) Pauly and Redisch (1973), and more recently by Rose-Ackerman (1986,1987) and Weisbrod (1988)¹. All of these works provide useful insight into the organization and operation of NFP hospitals. Yet, remarkably, no analyst has yet modeled the behavior of NFP hospitals in the presence of entry either by other NFP or for-profit hospitals, or does any of them model explicitly the market interaction between doctors and hospitals, or discuss equilibria that may emerge when both hospitals and physicians behave as strategic agents in the market.

Four distinct models of NFP hospitals deserve elaboration. Arrow (1963) argued that physicians seek to maximize their patients' health. He explains the hospital as an institutional guarantee to the client that the physician will ignore economic concerns - dealt with by the hospital - and make decisions only on the basis of what is best for the client.

Newhouse (1970) focused his attention on the decision making unit that determines the hospital's policy: the "administrator". Unlike most top executives, "administrators" of NFP hospitals do not maximize profits but maximize utility in a quantity-quality space. The hospital will produce a certain quantity at a certain quality so as to reach the highest feasible indifference curve of the "administrator" in this two dimensional space.

In a separate vein, a number of authors have attempted to understand the role of the hospital's medical staff in affecting the

behavior of the hospital. Most prominent in this literature has been the work of Pauly and Redisch (1973) and Pauly (1980). This literature holds that the NFP hospital exists primarily as a device to transfer economic profits from the hospital to its affiliated medical staff. The mechanisms by which this happens, and the allocation of those profits among the various members of the medical staff remain somewhat incomplete in this literature, but the primary idea is one of "doctor capture."

A separate and useful discussion of the NFP hospital appears in several works by Harris (1977,1979), assessing both the relationships between doctors and hospitals in resource utilization, and emphasizing the multi-product "job shop" nature of the hospital. Harris argues that hospitals incorporate two firms in one: Physicians are engaged in a non-cooperative game with the administration. Physicians maximize their own benefits not in cooperation with, but in spite of the administrative efforts to minimize costs.

In this paper, we seek to integrate much of this literature, including both the behavior of the hospital and its associated medical staff. This model incorporates concepts from Arrow, Newhouse, Rose-Ackerman, Weisbrod, Harris, Pauly, and also from the literature on monopolistic competition (Chamberlin, 1933). Thus, our model provides a comprehensive framework that links together a large number of ideas that have previously remained separated in the literature on the economics of medical care.

We start from Newhouse's intuition about the motivation of the hospital, and we re-interpret this process as a "political decision rule" that determines hospital behavior (section II). We then show

how unrestricted entry of new hospitals induces a stable equilibrium in the supply of hospital services (section III), and discuss the nature of this equilibrium (section IV). Given the equilibrium in supply of hospital services, we employ a model of "monopolistic competition" to explain the behavior of physicians in this market (sections V and VI).

We model physicians as independent firms that adhere to monopolistic pricing, due to asymmetric distribution of information between physicians and their clients. However, physicians must compete with their colleagues over their clients, which results in a basic market structure of "monopolistic competition".

Unlike Pauly and Redisch (1973), we attribute less than full success to physicians in their effort to use hospital resources to maximize their incomes. Instead, we model hospital decisions about output and quality as an ongoing political process. Thus, the physicians' maximizing problem is constrained by the equilibrium in the supply of hospital services. In particular, we specify physicians' cost schedules as a function of the equilibrium price and quality in the supply of hospital services. In parallel, we include physicians as an argument in the "political decision rule" that determines the supply schedule of hospital services (section VII).

We end up with a model that can be summarized in game theoretic terminology as a two stage sequential game. In the first stage, a "political decision rule" determines the supply schedule for hospital services. In the second stage, physicians make their decisions regarding the supply of physicians' services, given the equilibrium reached in the market for the supply of hospital services. To those

unfamiliar with game theoretic terminology, it may be worth emphasizing that this structure of the game forces the "political decision rule" to take into account, in the first stage, the rational behavior of the physicians in the second stage, since their behavior in the second stage will influence the final outcome or payoff for the "political decision makers."

II. A Utility Maximizing Director

We begin with a summary and re-interpretation of the model provided by Newhouse (1970), because, with proper interpretation, it incorporates many other models of hospital behavior, including a pure for-profit hospital.

Following Newhouse, consider a hospital directed by a sovereign "dictator" with a utility function in two characteristics of the hospital, say, Quantity (QN) and Quality (QL). Thus, $U=U(QN,QL)$. We will model the hospital as if it produces only one output (one type of care) and one quality, but in fact, the hospital's vector of outputs can include not only different types of treatment (as emphasized by Harris, 1977, 1979) but also different qualities of treatment of the same type. Thus, both QN and QL are vectors, possibly of large dimension. As will soon become apparent, the limitation to two dimensions eases the geometric portrayal of the equilibrium, but is not a fundamental feature of the model, and the role of the sovereign dictator can, at least in some circumstances, be further generalized.

The NFP hospital has several legal distinctions of importance for its behavior. Most prominently, it has no formal residual claimant, since it cannot have shareholders or owners.² While NFP hospitals may earn and accumulate profits, they must eventually disperse those

profits through activities of the organization, rather than as cash distributions to owners. We ignore the possibility of intertemporal accumulation of profits, and model this as a constraint that revenue of the firm equals costs, so that $P=AC$.

This organizational structure creates the essence of the Newhouse model of behavior of the NFP hospital. In a world where only the dimensions of (QN, QL) exist, the hospital "spends" its monopoly rents on those two "goods," by raising quality, lowering prices, or both.

The Lagrangian formulation of this problem shows:

$$(1a) L=U(QN, QL)+\lambda(P(QN, QL)\cdot QN-C(QN, QL))$$

Thus, the first-order conditions for the hospital's problem require:

$$(1b) U_{QN}+\lambda[P(QN, QL)\cdot(1+1/\eta)-C_{QN}]=0$$

$$(1c) U_{QL}+\lambda[P_{QL}\cdot QN-C_{QL}]=0$$

$$(1d) P=C/QN=AC,$$

where η is the own-price elasticity of demand confronting the hospital (holding QL constant). Using (1b), the equilibrium price solves readily:

$$P=(MC-U_{QN}/\lambda)\frac{\eta}{(1+\eta)}$$

where both MC and U depend both on QN and QL .³ This relationship also solves readily for $\eta = -P/(MC-P+U_{QN}/\lambda)$. Feldman and Dowd (1986) use this relationship to estimate η , although they maintain the strong assumption that $U_{QN}=0$ even in NFP hospitals. They estimate η in the range of -2 to -6 for various buyer groups from hospitals in the Minneapolis, MN area. However, if $U_{QN}>0$, they have overstated $|\eta|$, since they ignore this term in the complete definition of η . The authors separately estimate η directly for these same hospitals in the more traditional style using demand curves, and obtain estimates in the realm of $-.75$ to -1.1 .⁴ They attribute the difference to measurement

error in the price variable. The obvious alternative remains that utility from QN alters hospitals' behavior away from the pure profit-maximization solution, as directly predicted by our model.

If we solve equation (1b) and (1c) for λ , the interpretation would arise that the NFP hospital finds the optimum by setting equal the ratios of marginal utility to marginal net revenue on all dimensions of choice. In Eq. (1b), for example, marginal net revenue is $P(1+1/\eta)-C_{QN}$, as is $P_{QN}Q_N-C_{QL}$ in Eq. (1c). Thus marginal net revenues serves as "the price" to the NFP hospital when its directorate comes to solve its maximization problem.

In more intuitive terms, the hospital confronts a family of inverse demand (marginal value) curves for different qualities, as portrayed in Figure 1. At higher quality, patients' willingness to pay is higher at any quantity, so higher quality care leads to higher demand curves. The index i indicates increasingly higher qualities of care. At least past a given point, higher qualities also cost more to produce, so for every output, the average (and marginal) costs of production increase as quality increases. The hospital's demand curves arise from the mix of demands in the patient population being served.

** Figure 1 about here **

The NFP hospital must produce at a point where the demand curve crosses the AC curve.⁵ In general, this can occur at two points, one point, or no points. If the demand curve crosses the AC curve at two points, the utility maximizing decision maker in our model always picks the lower-right of the available points, since quality is held constant on any AC_i and D_i pair, and more QN is more desirable than less. If D_i cuts AC_i at only one point (tangency), this will always occur on the

left of the AC curve, since demand curves slope downward. If no intersection occurs, that quality level is not feasible to the hospital.

Figure 1 shows a subset of all possible qualities, with intersections qualities 1, 2, and (generically) i . The set of all possible such intersections forms the path EE, which represents market-feasible combinations of quality and quantity, given the NFP legal constraint that $P=AC$.⁶ The EE path slopes upwards at low levels of QL in Figure 1, but eventually, it will have a downward (backward bending) slope, so long as consumers have diminishing marginal valuation of quality and the average cost of producing quality increases with quality.

Which of these market-feasible points is most desirable to the sovereign hospital director? This becomes a standard problem in utility theory. The hospital director has a utility function in two goods, producing a set of indifference curves with slope $-U_{QN}/U_{QL}$, as in Figure 2. The EE curve from Figure 1 directly translates into a production possibilities frontier, labeled FF in Figure 2, that maps (QN, QL) pairs from the EE curve in Figure 1. FF is a monotonic transformation of EE, given ordinary relationships between quality and cost. The optimal point on the EE curve comes from the familiar tangency of the decision maker's indifference curves and the production possibilities frontier FF. As Figure 2 clearly shows, no sensible decision maker would willingly operate on the upward sloping portion of the FF curve (EE in Figure 1) because both quantity and quality would increase by a shift to higher quality. Thus, only the backward bending portion of the EE curve is relevant in Figure 1, comparable to the concave portion of FF in Figure 2.

** Figure 2 about here **

Generalizations of this to other models of hospital behavior (including Arrow, Pauly and others) come by considering how "real" hospital decisions are made. The crux of the NFP form of ownership -- the lack of shareholders to whom profits are dispersed -- means that any profits actually earned by the hospital become the common property of all persons who have legitimate access to them. The solution to this common property problem lies in political science, not economics. The decision rules established in the hospital provides various economic agents (administrator, trustees, nursing staff, medical staff, yes, even patients) with access to the hospital's profits through committees, formal organizational structure, and informal bargaining among parties. Models of the Congress may be more appropriate than models of the firm to understand this behavior.

A broad literature has emerged to suggest that equilibria in such settings may depend crucially on the rules of interaction (e.g., voting rules, agenda setting) or the organizational form, since without such rules, simple majority voting may lead to complicated cycling of group preferences (Shepsle 1979, McKelvey and Ordeshook 1984, Banks 1985). We do not seek here to resolve the question about how equilibria in such settings are achieved, but rather, we assume that some decision process has emerged that allows the organization to exhibit stable behavior. So long as the political power of each agent in the hospital remains constant, the "preferences" of such an organization will remain stable. In other words, the hospital will act "as if" it had a single decision maker directing its decisions, similar to Newhouse's sovereign hospital administrator.

Naturally, to accommodate multiple economic agents in such a political process, the hospital's decision rule must have more than two arguments. Thus, more generally, $U=U(QA, QB, \dots, QL, QM, QN, \dots, QZ)$ describes the sovereign director's apparent utility function. Graphing such a function offers a more complicated artistic problem than the two dimensional utility function we first discussed, but otherwise, the general ideas remain unchanged. Thus, for expositional simplicity, we will continue to use the two dimensional utility function first proffered by Newhouse.

From this approach, most previous models of hospital behavior appear as corner solutions in what we view as a broader bargaining game. In the classic Newhouse model, the administrator wins it all. In the Pauly-Redisch model, the medical staff wins it all (although the doctors confront a similar problem of allocation within their own ranks that Pauly and Redisch discuss hardly at all). Other models have suggested that the hospital uses its profits to pay excess wages to the nursing staff, or, conversely, exploits the nursing staff monopsonistically [Booten and Lane, 1985], or that the primary beneficiaries of the NFP structure are patients [Arrow, 1963]. In our model, each of these models could represent the goals of the participating economic agents, but we do not expect corner solutions in general. Rather, we anticipate that each of these agents will have only some of their wants satisfied, and that profits will be divided among agents in any equilibrium that may emerge.

If the political bargaining rules change for some reason, then the apparent utility function will shift. For example, if a labor union is certified to bargain for the nurses, their power could

plausibly increase. A court ruling concerning the composition of medical staffs would alter the power of physicians. Some changes in the form of hospital reimbursement could shift political power within the organization, while others would merely act as a shift in the overall budget, affecting all parties equally. In any event, if the power does shift, then the apparent utility function will shift. The "phantom" sovereign administrator will appear to have changed her mind -- the hospital decision rule will change. In much of what we discuss below, external changes will not obviously alter a hospital's decision rule, so unless noted otherwise, we will presume that we can act as if the hospital had a stable decision rule, and proceed with a Newhouse-like utility function to represent the preferences of the phantom decision maker.

III. Not-For-Profit Hospitals with Entry.

The model described above and portrayed in Figures 1 and 2 remains valid if and only if the hospital has a monopoly, or if the quality and price of all other hospitals in the market remains unchanged. Under those circumstances, we can meaningfully talk about a stable set of demand curves for various levels of quality of the hospital. However, entry by a new hospital, or changes in scale or quality of another hospital, alter the entire market structure.

a. New Entrants and Existing Hospitals' Demand Structure

Consider an established monopolist (Hospital A) who faces a new entrant (Hospital B), and consider the choices confronting the new entrant B. If B enters at some arbitrary scale and quality, some of A's patients will end up in B. The entry by B will shift inward the entire family of demand curves for A from D_1 to D_1' (dashed line demand

curves in Figure 3), although not uniformly, unless B enters with exactly the same quality as A. For example, if B enters at a higher quality than A, the higher-quality demand curves confronting A will shift inward more than the lower quality demand curves, and conversely. Assuming that cost curves do not shift, the EE curve for the hospital shifts inward in Figure 3 to E'E' and the FF production possibility curve in Figure 4 also shifts inward to F'F'.

** Figure 3 about here **

** Figure 4 [a & b] about here **

Figure 4a shows the effect on A when B enters at higher quality than A. A's optimal response leads to lower quality and quantity. Similarly, Figure 4b shows the effect on A's choices when B enters at lower quality. The new output of A is both lower in quality and in quantity, but the decline in quality is not as large as in Figure 4a. It is easy to see that A's quality could actually increase in some settings.

When contemplating entry into this market, hospital B's directors can anticipate changes in the style of A, and plan accordingly. For any possible quality they might produce in B, B's directors can anticipate the reaction of A and hence the location of the family of demand curves confronting B **after** A has made its optimal move. These demand curves form the basis of B's entry decision, and the shape of its own EE curve. By tracing out all possible qualities, equivalent reactions of A, and final demand curves confronting B, the directors of B can pick (using their own utility function) the optimal quality with which to enter. Thus, B and A should have a stable division of the market, so long as B and A can properly anticipate each other's reactions. This is a standard problem in duopoly, and the reaction functions of A and B can be constructed in the usual fashion to derive the eventual equilibrium.

In any market where factor supplies have constant prices, the AC curves for the hospital will not shift when the entrant appears. Our previous discussion made this assumption. However, if factor supply curves slope upward, then as entrants increase aggregate factor demand, factor prices will rise, and hence the AC curves for the incumbent hospital will also rise for all levels of QL. This further exaggerates the inward shift of the EE curve in Figure 3 arising from the shifts in patient demand.

b. Economic Limits on Entry

Entry will not continue unabated forever in this market. Hospitals B, C, ... all need capital to enter. Without equity markets available to them, either bond markets or donors must supply needed capital. Both markets require that the hospital ultimately prove capable of survival before providing such capital. In the bond market, the hospital has the advantage of using bonds paying interest that is not taxable to the investor, thus lowering the cost of capital somewhat, but this only changes the equilibrium level of output.⁷ In the market for donors, competition across hospitals (and other charities) will constrain the behavior of the hospital similarly (Rose-Ackerman, 1987).

To understand how entry is ultimately constrained, consider some quality QL_j , such that the demand curve confronting hospital B at QL_j never crosses the relevant average cost curve AC_j . This quality choice is not feasible for hospital B. If some cutoff quality QL^{\max} exists, above which demand never intersects $AC(QL^{\max})$, then the EE curve terminates at that quality. If no quality exists at which the demand curve intersects the average cost curve for that quality, then hospital B can never enter.

The role of entry, and indeed, the lack of any consideration of entry in the Newhouse (and subsequent) models appears most notably in their proclivity to draw demand curves crossing AC curves at more than one point. For this to hold, the hospital **must** have monopoly power in the traditional sense (that the demand curve is downward sloping) and also in the monopolistic competition sense (that a combination of price and output is feasible that would create economic profits). Persistence of a situation where demand curves intersect AC curves twice would imply some sort of barriers of entry.

If entry can take place continuously and smoothly, then eventually each hospital will confront a monopolistic competition environment at any possible quality it could consider, with demand curves at each quality exactly tangent to the associated AC curve. This also creates an EE path, just as in the previously discussed monopoly case, except that excess capacity must exist, with average costs declining at equilibrium for all qualities. By contrast, in the case of limited entry, the hospital **could** operate in the realm of increasing average costs, depending on whether or not the demand curves intersected AC on the right or left of AC_{min} .⁸

As we observe hospital behavior, it would appear that individual hospitals have considerable flexibility in altering the vector of qualities of output. Changes in staffing mix (e.g., from RNs to LPNs) will alter overall quality. Dedication of a small unit of beds to a specialized intensive care unit (with more intense monitoring equipment, specialized intensive care nurses, etc.) also augments quality for a subset of the hospital capacity directly, and it also increases the quality of "normal" beds by providing the standby "emergency" capability for more intensive care, if needed. Hospitals would appear

to have considerable latitude in adjusting the size and intensity of such units. Provision of specialized diagnostic or treatment services (MRI, lithotripsy, etc.) also offers ways to upgrade "quality," albeit only for subsets of all possible patients.

Hospitals can also augment quality in the "amenities" dimension of care, through changes in such obvious characteristics as food service, decor, quality of television in patient rooms, etc. The hospital can also shift capacity from lower quality "semi-private" rooms to purely private rooms, simply by removing a bed from the room and adding extra chairs for visitors. In addition, entire "wings" of hospitals can be shifted from semi-private to private rooms, much as hotels commonly offer "executive" floors with special decor and services. Thus, it would appear that hospitals have considerable flexibility in this dimension of quality as well.

In general, we would expect that the capability of smooth and nearly continuous adjustments in the mix of qualities of care would lead to few unexploited opportunities for entry in every niche of specialized quality. The sole exception, of course, would come if overall capacity of the hospital market were externally constrained, e.g., through strictly enforced hospital planning laws. We discuss next the role of legal limits on entry.

c. Legal Limits on Entry

Government planners have persistently confronted questions of "excess capacity" in the hospital sector, and have sought through various regulatory interventions to limit the entry of new capital into the industry. The most recent of these rules, the hospital "Certificate of Need" (CON) laws provide a good example of this style

of regulation. State governments (under Federal authorization and financial support) established CON agencies that in turn established region-by-region targets for the number of hospital beds, and commonly for any large-cost capital investment.⁹

In concept, such CON laws could actually improve the efficiency of a market, by limiting entry just to the point where each hospital produced at its minimum average cost. To see how this might happen, note in Figure 1 that the EE curve for a single-hospital monopolist passes through points on the right hand side of AC, where $MC > AC$.

As we have also discussed previously, with unconstrained entry, we can expect that the EE curve will eventually contain only points of tangency of demand curves to AC curves, each at points where $AC > MC$, with classic Chamberlinean excess capacity. It follows that intermediate amounts of entry would create EE curves between these two extremes, and thus that some entry constraints might actually cause the EE curve to pass uniquely through points of minimum AC for all quality levels. In order to achieve this, of course, the CON regulator would necessarily have to dictate the rate of entry at each feasible quality. The informational requirements to achieve this appear considerable, but in concept, regulation could achieve this outcome.

If the monopolist's EE curve passes only through the left side of AC curves, a technical possibility, then any entry will only increase AC of the existing hospital. This would correspond to the usual "natural monopoly" where AC curves declined for the entire range of relevant output, although in this case, entry could still occur in the absence of restrictions.

In actual practice, CON laws have proven less effective than might have been hoped. Several studies suggest that CON laws have reduced the number of hospital beds below what might have otherwise occurred, but that overall hospital costs were unaffected by such laws (Salkever and Bice, 1976; Sloan and Steinwald, 1980; Joskow, 1981).

Such CON regulators may also reduce social welfare, even if well-intentioned. For example, if they overly constrain entry, so that the EE curve consists of points to the right of AC_{min} for incumbent hospitals, then AC will be higher than if more entry were allowed. It is even possible, of course, for the CON process to limit entry so much that AC exceeds that which would emerge with unconstrained entry.

The guiding rule for any well-intentioned regulator would be whether or not its subject hospital industry displayed increasing or decreasing average costs. In general, if the AC curve slopes upward at equilibrium output for the hospital, then the CON regulators have been overly vigorous in restraining entry, and conversely.

IV. Discussion of the Hospital Market Equilibrium

Several pertinent issues emerge from this analysis. We now discuss some of them that seem most relevant to the main themes of this article.

a. SRAC vs LRAC.

First, consider the distinction between long run and short run average costs. This distinction matters in the NFP analysis because the legally binding constraint that $P=AC$ must hold whether or not the hospital operates at its most efficient capital stock or at some less efficient level. However, the NFP hospital will have incentives to operate on the LRAC curve whenever possible, even if it has a monopoly in the market. Consider a hospital as in Figure 5, showing only a single quality of care. (An entire family of such curves

exists for the various possible qualities of care.) If the firm has a capital stock that provides it with $SRAC_1$, the legal constraint that $P=AC$ will allow it to produce only at QN^1 . However, by investing in more capital, it can move to $SRAC_2$, the $SRAC$ curve touching the $LRAC$ curve at QN^2 , the hospital's ultimately desired output level. Since QN augments utility of the decision maker, the hospital will eventually move to $SRAC_2$, i.e., producing on the $LRAC$ curve.

** Figure 5 about here **

Competitive entry will also force hospitals to move to the $LRAC$ curve tangency, rather than allowing it to persist with an inefficient level of capital. To see this, note that if a potential entrant hospital (with access to the same technology) confronted an incumbent hospital producing on a $SRAC$ curve at points other than on the $LRAC$ envelope, it could enter, produce the same quality at a lower price, and draw patients away from the inefficient incumbent.

b. Mixed Markets with NFP and For-Profit Hospitals

Some regions in the United States have NFP and for-profit hospitals operating in competition with one another. Several analysts have sought to demonstrate what differences, if any, occur in their operation, including questions of efficiency of operation, willingness to treat indigent patients, propensities to offer various services, etc.

As we discussed in section IIIb, with unconstrained entry, every hospital in the market will find itself operating in a monopolistic competition tangency of demand curves and AC curves. This holds whether the hospital is NFP or for-profit. In this setting, the for-profit entrant will also have to select a quality at which to operate, but with unconstrained entry, the opportunities for profit would be the same for all levels of quality. In general, except for differences in

cost of production due to legal differences, we would not expect to see substantial differences in the behavior of for profit and NFP hospitals in such a setting.

NFP hospitals operate at an advantage in this setting because of the preferential tax treatment (thereby avoiding income taxes, property taxes, etc.), and because of the lower cost of capital (through tax-free bonds and because of tax-subsidized donations). Despite these advantages, studies comparing costs of hospitals find little difference in their outcomes. (Becker and Sloan, 1984, Watt et al, 1986). Some authors have argued that the inherent inefficiency of NFP hospitals, where property rights are not complete, offsets any cost advantages these organizations could otherwise have, but another obvious difference also exists: the NFP hospital may be operating at a different "quality" because of the preferences of the hospital director.¹⁰

The for-profit monopolist operates as a special case of the NFP model described by Eq. (1), with $U_{QN}=U_{QL}=0$. Thus, the for-profit hospital concerns itself only with profits, and the constraint in Eq. (1) becomes the objective function. The first order conditions in (1b) and (1c) apply directly with this modification. Thus, the for-profit hospital will produce lower QN and QL than an NFP would with access to the same technology and confronting the same demand curves, and would charge a higher price.¹¹

This analysis emphasizes the importance of controlling for "quality" in comparisons across hospitals. Without such careful control, the costs of one hospital could appear higher or lower than those of the other without any implications for technical efficiency of production.

The presence of potential entry also constrains the behavior of the NFP sector. For example, if the preferences of a NFP "director"

caused the hospital to produce only at very high quality, then entry by a competitor hospital at lower quality would cause substantial loss of patients by the incumbent (see Figure 4b). Obviously, if the NFP incumbent is protected by legal limitations on entry, then the constraints imposed by competition would not appear. Similarly, if entry is blocked but the incumbent is a for-profit hospital, we would expect the hospital to price in traditional profit-maximizing fashion, i.e., where $MR=MC$ for the demand curve confronting the hospital.

V. Physicians in Monopolistic Competitive Environment.

So far, we discussed the decision of the NFP hospital with respect to its production function as if the only thing that distinguished it from other firms was the lack of shareholders to whom profits are dispersed. However, NFP hospitals differ from other firms in yet another crucial feature: their dependence on physicians who, in many respects, behave as independent firms in this environment. We now turn to explore this element of the production of medical care.

a. Physicians in Monopolistic Competition.

To describe the supply side of the health industry, we use a model of a "monopolistic competitive" market.¹² Assume (1) a set of N physician-firms that produce similar but not identical products; (2) prices are a function $p_i(q_i; Q_{-i})$ where q_i is the quantity produced by physician i , and Q_{-i} is the quantity produced by all firms except i ; (3) U shaped average cost schedules for the firms. (4) free entry in individual markets, possibly constrained by entry barriers in the aggregate that allow firms to realize positive rents; and (5) price search by consumers is too limited to allow the market to reach a purely competitive solution.

The first three assumptions are basic assumptions of the monopolistic competition model. The first is justified by the heterogeneous character of the supply of "medical care".¹³ The second and third are nonproblematic in this context. The fourth is discussed in a lengthy endnote.¹⁴ An adequate discussion of the fifth assumption is beyond the scope of this paper. However, in subsection (Vb) below we discuss this assumption and explain how our analysis is consistent with some level of consumers' price search.¹⁵

Assume a set of physicians $N=\{1,2,\dots,n\}$, producing similar but not identical services. In a "monopolistic competitive" market, each physician in the market faces the following maximization problem: $\text{MAX } p_i(q_i; Q_{-i})q_i - c_i(q_i)$, where q_i is the quantity that physician i produces, and Q_{-i} is the amount produced by all physicians except i .

Let $D^{-1}=K-bQ$, represent the market inverse demand curve, where Q is the aggregate demand from all buyers, which in equilibrium equals aggregate quantity supplied, s.t. $Q^* = \sum_{i=1}^N q_i^*$

Chamberlin's model of "monopolistic competition" [1933] rests on the assumption that both demand and cost curves for the ("different" heterogeneous) products be uniform across firms (physicians). In our notations this implies that $Q=n \circ q_i$. This assumption is hard to justify, and unnecessary. While this assumption can be easily relaxed, for expositional purposes we only allow the quantity produced in the market to depend on i 's production and $(n-1)$ identical physician-firms which we will denote by $j \neq i$, s.t.:

$$Q^* = \left[\sum_{j \neq i}^N q_j^* \right] + q_i^*$$

Let $C_i(q_i) = \alpha_i \circ q_i^3 - \beta_i \circ q_i^2 + \gamma_i \circ q_i + \delta_i$. The third power assures the U shape of the MC and AC curves. $-\beta \circ q_i^2$ shifts the minima of the MC and

AC curves to the positive range of Q. Finally, $\gamma \circ q_i$ guarantees that MC never falls to the negative range of the price axis, and δ_i represents the fixed costs. Firm j faces a similar cost schedule:

$C_j(q_j) = \alpha_j \circ q_j^3 - \beta_j \circ q_j^2 + \gamma_j \circ q_j$, but may be distinguished from firm i by the parameters of the cost function, since $\alpha_j \leq \alpha_i$, $\beta_j \leq \beta_i$, $\gamma_j \leq \gamma_i$, $\delta_j \leq \delta_i$.

Chamberlin modeled the dynamics of entry by parallel shifts of the demand curve inwards [Chamberlin 1933, pp. 83-84]. In a substantial deviation from Chamberlin, we model the dynamics of entry by **rotating** the firm's demand curve inwards s.t. K remains fixed. If we assumed no search, each firm receives a random portion of the aggregate market demand, and the firm's demand curve is a scaled-down version of the market demand curve, with intercept K on the price axis. Adding new firms to the market causes the demand facing each existing firm to fall at all prices, causing the inward rotation. This modification increases substantially the explanatory power of the model, and we can readily generalize it (see section Vb) to allow for some price search.

Under these assumptions, physician-firms face the following maximization problem:

$$(2) \text{ MAX } P(Q) \circ q_i - c_i(q_i), \text{ or:}$$

$$(2a) \text{ MAX } \left[K - b \circ [(n-1) \circ q_j + q_i] \right] \circ q_i - c_i(q_i)^{16}$$

The first order condition [by substitutions and differentiation] is:

$$(2b) \left[K - b \circ [(n-1) \circ q_j + 2q_i] \right] - q_i \circ \left[b(n-1) \circ \frac{dq_j}{dq_i} \right] - MC = 0, \text{ or}$$

$$(2c) \left[K - b \circ [(n-1) \circ q_j + 2q_i] \right] - q_i \circ \left[b(n-1) \circ \frac{dq_j}{dq_i} \right] - 3\alpha \circ q_i^2 + 2\beta \circ q_i - \gamma = 0, \text{ or}$$

$$(2d) \left[K - b \circ [(n-1) \circ q_j] \right] - q_i \circ \left[b(n-1) \circ \frac{dq_j}{dq_i} \right] - 3\alpha \circ q_i^2 + (2\beta - 2b) \circ q_i - \gamma = 0, \text{ or}$$

$$(2e) K - b \circ [(n-1) \circ [q_j + q_i \circ \frac{dq_j}{dq_i}]] - 3\alpha \circ q_i^2 + (2\beta - 2b) \circ q_i - \gamma = 0.$$

The second order condition requires that $\Pi''(q_i) < 0$.

$\frac{dq_j}{dq_i}$ is of special interest here. It represents the change in quantity produced by any other physician j that will result from any marginal change in the quantity produced by physician i . Another interpretation that has recently gained some attention (e.g. Bresnahan 1981) interprets $\frac{dq_j}{dq_i}$ as a conjecture (or belief), say ψ_{ij} , that physician i has about the change in quantity produced by physician j that will result from any marginal change in the quantity produced by physician i . It is now easy to see how Chamberlin's assumption that $Q = \sum_{i=1}^N q_i$, coupled with what is known the "Cournot conjecture" that $\frac{dq_j}{dq_i} = 0$, simplifies the analysis, since introducing these two assumptions reduces (2e) above, to:

$$(2f) \quad K - b \cdot [(n-1) \cdot q_j - 3\alpha \cdot q_i^2 + (2\beta - 2b) \cdot q_i] - \gamma = 0,$$

but since $q_j = q_i$ we can write:

$$(2g) \quad K - b \cdot (n+1) \cdot q_i - 3\alpha \cdot q_i^2 + 2\beta \cdot q_i - \gamma = 0,$$

which can be easily solved for q_i .

Figures 6a and 6b illustrate the crucial features of this model. As n increases, the demand curve that each physician faces rotates inwards s.t. the absolute value of the price elasticity of demand faced by the physician remains constant.

* FIGURE 6 about here *

For simplicity, we have presumed that the physician firm produces only one output (q_i). The market demand curve appears as D_1 , with its corresponding marginal revenue curve MR_1 . In Figure 6a, if (for whatever reason) only a single physician-firm operates in this market, then the market demand curve and the demand curve facing the firm are identical, and price is set at P_1^* using the usual rules that the firm sets output s.t. $MR=MC$, i.e., at q_1^* .

Suppose now that a second firm enters the market. Given the limited amount of search, the two firms will end up randomly dividing the patients between them, so that each confronts a demand curve like D_2 , (equivalent to MR_1 in this case) with its corresponding MR_2 curve. Each firm sets the [optimal] price at P_2^* , lower than P_1^* , because MC of production are lower at q_2^* than at q_1^* . However, both firms make positive profits, thus making entry still desirable. If the economic returns from this market exceed those from other markets, and if no obvious barriers to entry exist, then this market is not in equilibrium with two firms.

Suppose now that six more firms enter (still presuming no patient search), with patients randomly distributed across the eight firms, making the relevant demand curve D_8 . In this setting, monopolistically competitive pricing leads to the classical tangency equilibrium. If each firm attempted to reach the pure monopoly price and output, average cost would exceed price, and no firm could survive. Thus, each takes advantage of the declining average costs in this region of output, and seeks the point of tangency where AC and demand curves are tangent, yielding the monopolistic competition equilibrium price of P_8 with output per firm of q_8^* . In this setting, no firm has the incentive to enter or exit, so this represents a stable equilibrium. Since we believe, as we mentioned above, that entry is restricted at the aggregate, we expect that at equilibrium rents will be positive, and (in this example) the equilibrium number of physician-firms is $n^* < 8$.

As long as the MC and the MR curves intersect where MC is upwards sloping, P^* is decreasing in n . However, due to positive rents, entry will continue. Eventually, the intersection between the MC and the MR

curves may shift to the range where MC is downwards sloping. At that point, P^* will start increasing in n . With our cubic function for total cost, minimum AC occurs at $q_{min}=\beta/2\alpha$, and minimum MC occurs where $q=\beta/3\alpha$. Thus, MC is falling for all $q<(2/3)q_{min}$. Thus, despite the common practice of pairing U-shaped AC curves with **only** upward sloping MC curves, we rely on the presence of declining MC for a wide range of output.

In Figure 6a, between $n=2$ and $n=8$ firms, prices rise with n . This occurs because MR intersects MC in the realm where MC falls with output. Thus, the lower the output of each firm, the higher the MC which in turn leads to higher prices. Figure 6a shows this for linear demand curves. It is easy to demonstrate the underlying logic when demand curves have constant elasticities. As long as search is limited enough, entry of a new firm leaves η_i unchanged for each firm, but the output of each firm, q_i , falls. Thus at the range where MC is declining the monopoly price $P=MC \cdot (\eta/(1+\eta))$ must rise.

Figure 6a shows the situation where entry first causes prices to fall, and then to increase in n . Obviously, as prices rise, the total quantity produced by the market will start to decrease in n . Figure 6b shows a situation where the "competitive monopolistic equilibrium" is reached at $n=4$ firms, where prices are still decreasing in n .

Some empirical evidence exists for a positive relationship between physicians per capita and prices of physicians' services. Some authors interpreted such findings as evidence for induced demand, including Evans' [1974] "target income" model. Pauly and Satterthwaite (1981) estimated the relationship between physicians per capita and prices as a test of Evans' model. They admit that they can not reject the

"target income" (or induced demand) hypothesis, but argue that the empirical evidence can be interpreted, alternatively, as a support for their "increasing monopoly" model. Their model rests on the assumption that search costs increase in n . Under this assumption, as the number of physicians per capita increases, the level of price search in the market decreases, which results in smaller absolute values of price elasticities of demand. Again, by the standard monopolistic pricing model, firms maximize profits when $P=MC \cdot (\eta/(1+\eta))$. By assumption, search costs increase in n and an increase in search costs results in a decrease of $|\eta|$. Since monopolistic pricing requires that $|\eta|>1$, unless MC fell rapidly as output fell the Pauly-Satterthwaite result follows directly.

Our model provides an alternative theoretical explanation for the positive relationship between the ratio of physicians per capita and prices of physicians' services -- it is a natural outcome of the monopolistic competition between physicians in a market where the level of search is constant and is not high enough to allow a competitive market equilibrium. In contrast to the "increasing monopoly" model suggested by Pauly and Satterthwaite [1981], we expect expect prices to rise with n , not because of increasing costs of price search, but simply because of the inwards rotation of the demand curve, due to entry in a monopolistically competitive industry, which induces tangencies with AC curves at higher prices. We get this result holding search costs [and thus search] constant.¹⁷

In our model, if $MR=MC$ where marginal costs are decreasing (and $|\eta|>1$), as in Figure 6a, then entry of new physicians will result in price increases. However, if $MR=MC$ where MC is increasing, then entry

will result in lower prices. The empirical results reported by Pauly and Satterthwaite (1981, table 3, p. 503) are consistent only with the situation portrayed in Figure 6a. Our model is also consistent with the recent history of the U.S. medical economy, where physician prices have increased faster than inflation, but physicians' real earnings have remained essentially flat for the last decade. The substantial entry by new medical school graduates would create the increases in physicians per capita that would produce all of these phenomena. If this is the case, and if for any reason we wish to reduce the price of physicians' services, then we should either limit entry or take steps to enhance price search by consumers.

We anticipate settings like Figure 6a to occur in medical markets in the U.S. due to the extensive insurance held by most Americans. Such insurance makes all demand curves intrinsically less elastic (Phelps and Newhouse, 1974), and also reduces the incentive of consumers to search. Thus, situations where demand curves are very steep are more likely to produce equilibria in the realm where MC is decreasing in q and hence where prices increase in n .

b. Price Search

We can now relax the assumption of no search. Search by consumers across firms increases the absolute value of the slope of the demand curve confronting each physician-firm, while holding constant the composition of demand at the market level (Sadanand and Wilde, 1982). However, the discussion surrounding Figures 6a and 6b remains valid even if we make the demand curves confronting each firm more elastic.

So long as the level of search does not admit purely competitive markets, the results we discussed earlier and illustrated in Figure 6

continue to hold.¹⁸ These results persist so long as the amount of search remains constant as n changes.

In fact, empirical evidence is inconsistent with the assumption of "no search". Substantial empirical evidence exists to support the hypothesis that $0 < |\eta| < 1$, at the market level (Phelps and Newhouse 1972, 1974; Manning et al. 1986). Monopolistic pricing requires that, for the firm, $|\eta| > 1$, otherwise the second order conditions for profit maximization are violated. McCarthy (1985) estimates $\eta = -3$ for primary care physicians. The only way to reconcile such small observed market-level price elasticities of demand with monopolistic pricing and with the empirical evidence that is consistent with monopolistic pricing is to accept the notion that some price search occurs. Empirical evidence shows that consumers do engage in some levels of price search (Olsen, et al. 1976). This search, even though incomplete, provides a mechanism to distribute patients across firms, as our model requires.

c. The Role of the Hospital

We can now investigate one important interaction between hospitals and doctors: hospitals can reduce the costs of production for doctors by providing resources to the doctor at below-market prices, e.g., by "spending" hospital monopoly rents in this fashion. Consider again the cost function of physician-firms. Let $MC_i = MC_i(q_i, w_i, QN_h^*, QL_h^*)$, where q_i is the quantity produced by physician i , w_i is cost vector of the vector of input factors needed for production (physician's labor, investment in education, etc.), QN_h^* is the equilibrium quantity of each service that the hospital provides, and QL_h^* is the equilibrium quality that the hospital provides. We

discussed in the previous sections how equilibria in QN_h^* , and QL_h^* are reached. We now show how physician costs depend on the equilibrium quantity and quality produced by the hospital.

Figure 7 shows a physician market in several alternative situations. Two demand curves are shown, for n and $n+1$ firms. Suppose the "unsubsidized" costs of each of the n physician firms are as shown by AC_1 . In this case, the n firms would price at the standard monopolistic competition tangency, each selling q_1 at a price P_1 . Suppose now that the hospital provides some resources to the physician firm at below-market cost. If the hospital does so to the extent that costs fall to AC_3 , then another entrant will appear, and no physician will be better off since they will (again) price at a tangency (q_2, P_2). However, there will always exist some intermediate provision of resources that provides some cost reductions, say to AC_2 , but not sufficiently so as to provoke entry. At this level of hospital-provided subsidy, the optimal quantity (where $MR=MC$) leads to the same initial output and price (q_1, P_1), but each physician now makes monopoly rents. Obviously, AC_2 was selected to produce the original output and price, but any other AC curve lying between AC_1 and AC_3 would produce increased physician-firm profits without inducing entry, with the possibility that P^* will rise or fall, as the AC curve lays above or below AC_2 .

* FIGURE 7 about here *

Hospitals' cost-lowering activity, even to conditions like AC_3 could also benefit the existing medical staff, if doctors could somehow block entry from outside. The hospital provides a convenient vehicle to accomplish this, it turns out, through one of several mechanisms.

Suppose first that the hospital simply refused to admit any new physicians to its medical staff. This "closed staff" arrangement was actually not uncommon in the past. It represents a blatant barrier to entry with all of the usual consequences. This arrangement obviously matters more in the types of medical practice where access to the hospital has more effect on the physician-firm's cost structure. The conduct of complicated surgery within the hospital represents an obvious case where the hospital provides lower cost production than the physician-firm could produce alone. By contrast, such specialties as dermatology seldom use the hospital, and the hospital would be generally irrelevant to entry in such specialty fields.

Finally, hospitals could act as a coordinating agent for physician-firms by installing specialized equipment (e.g., special surgical equipment), but only at a capacity sufficient to support the incumbent firms. At this point, the entry-limiting aspects of the hospital's activity depend in part on hospital procedures and rules for allocating access to scarce resources. Assignment of operating room time, for example, might be made in a committee of the surgical staff in such a way as to effectively preclude entry.

Arrow explained the NFP hospital as a device to disassociate physicians from their economic incentives. While this may hold in part, we point out that the hospital can also benefit the doctors in three separate ways: First, physicians may benefit from expected economies of scale on technology and labor. Second, the hospital can, under some circumstances, facilitate cartel behavior for the physicians by creating subtle restrictions on entry.

Third, literature on the problem of "collective action" stresses that an organization can sustain cooperation among non-cooperative players, involved in a problem of collective action, if it is able to compensate its members for their cooperation by providing other (transferable) benefits (Olson, 1971).¹⁹ As most hospitals are NFP organizations, they can provide a mix of services and prices, as discussed above, to allow physicians to collect additional rent.

VI. Physicians and Hospitals: How Do Physicians Choose?

Now that we have discussed the supply of hospital and physician services separately, we are ready to combine the two and study their interaction in the production of medical care. We start by describing the second stage of this game when hospitals have already picked their equilibrium production and quality levels and physicians have to choose in which hospital to work. In the next section we conclude the analysis by folding back and describing how hospitals, anticipating physicians' behavior, choose their equilibrium quantities and qualities of production.

a. The Case of Only One Hospital in Town

Assume that each town has only one hospital. Pauly and Redisch (1973), assuming that physicians act cooperatively to maximize average income, found that if physicians control entry to the hospital ("closed staff"), then the maximizing solution occurs at the intersection of the marginal revenue product curve and net average revenue product curve. Relaxing the assumption of control over entry ("open staff") will alter the equilibrium to the intersection of the net average revenue product and the marginal supply price of physicians' services. Physicians in "open staff" settings will defect to build new hospitals

where they will hope to establish a "close staff" setting that would guarantee them a higher income. This provides an explanation of duplication in the market.

Our model provides an analogous result without the restrictive cooperation assumption. Physician entry will continue until rents are equalized across local markets. However, physicians could start leaving the hospital before this occurs, and establish a new hospital in a new town where higher-than-average rents can be realized.

The model is easily generalized to provide a framework of analysis for cross geographical equilibria. As long as there exist "towns" where unusually large rents are realizable physicians will price monopolistically and obtain net rents. As new hospitals enter the market, they must allow their medical staff to realize positive rents. This process will continue until all opportunities for positive rents are eroded in all markets. However, if barriers limit entry at the aggregate, then in equilibrium, holding everything else constant, rents will be similar across geographically separated markets.

b. More than One Hospital in Each Town: The General Case.

A multitude of hospitals leaves the physician with a set of alternatives $(Q_N^*, Q_L^*) = \{(Q_{N_A}^*, Q_{L_A}^*), (Q_{N_B}^*, Q_{L_B}^*), (Q_{N_C}^*, Q_{L_C}^*), \dots, (Q_{N_M}^*, Q_{L_M}^*)\}$, where A, B, C... label the different hospitals as in sections III and IV above, and each ordered pair $(Q_{N_j}^*, Q_{L_j}^*)$ is simply the equilibrium choice of hospital j regarding prices and quantities of production. Physicians will chose the hospital[s] that maximize[s] their utility. Thus, if we denote the choice as $C_i \{Q_N^*, Q_L^*\}$, each physician will select $C_i \{Q_N^*, Q_L^*\}$, such that:

$$C_i \{Q_N^*, Q_L^*\} = \max \{u_i(Q_{N_A}^*, Q_{L_A}^*), u_i(Q_{N_B}^*, Q_{L_B}^*), u_i(Q_{N_C}^*, Q_{L_C}^*), \dots, u_i(Q_{N_M}^*, Q_{L_M}^*)\}$$

where $u_i(Q_{N_j}^*, Q_{L_j}^*)$ is simply the utility that physician i gets from

practicing in hospital j , that produces quantity QN_j^* , at quality QL_j^* . Physicians need not choose to work in only a one hospital. Physicians often choose a "mix" of hospitals between which they share their time of practice. Physicians may find that they have to work in more than one hospital to produce their optimal output.

c. Physicians with Complex Preferences

If all physicians were indifferent between all geographic locations, economic rents should equalize across geographic locations (Newhouse et al. 1982). Some localities, however are more attractive for physicians for other reasons. In such locations, we expect lower rents for physicians, using a straightforward compensating differential approach. The effect of this on the hospital emerges most readily when we consider a three-good utility function for the hospital, so that $U=U(QN,QL,QT)$ where QT is a transfer to doctors. In two otherwise-identical hospitals, the equilibrium level of QT is smaller for a community with higher amenities for doctors. This reduces the importance of QT in the utility space, with the effect of shifting out the EE and FF curves in (QN,QL) space. Thus, other things equal, QN and QL , and hence average hospital costs, will all be higher in such communities than in "less desirable" communities.

If physicians have geographic preferences for cities like New York or Boston, over, say, New Haven, then rents for physicians should be higher in New Haven. Thus hospitals in New York will produce more and at a higher price, with physicians receiving lower average rents. In the long run, physicians who prefer money to metropolitan lives will end up in New Haven, while physicians who prefer the cosmopolitan life will chose New York. This extra complexity can easily be

incorporated into the framework outlined above. Instead of

$$C_i\{QN^*, QL^*\} = \max\{u_i(QN_A^*, QL_A^*), u_i(QN_B^*, QL_B^*), u_i(QN_C^*, QL_C^*), \dots\}$$

we will now have:

$$C_i\{QN^*, QL^*\} = \max\{u_i(QN_A^*, QL_A^*, L_A^*), u_i(QN_B^*, QL_B^*, L_B^*), u_i(QN_C^*, QL_C^*, L_C^*), \dots\}$$

where $u_i(QN_j^*, QL_j^*)$ is simply the utility that physician i gets from practicing in hospital j , that produces quantity QN_j^* , at quality QL_j^* in location L_j^* . This allows us to generalize our analysis to multi-dimensional utility spaces, allowing physicians to have preferences over localities, and, in each locality, over quality and rents. In turn, each hospital, in its decision on the levels of quality and quantity of production, must take into consideration the locational choices of physicians, as well as their preferences over the mix of quantity and quality. Thus, we expect differential rents in different localities and in different hospitals, due to preferences physicians over other "goods" such as attractiveness of the locality, quality and prestige of the hospital, etc. The same ideas apply to physicians' "practice style" (Boardman et al., 1983).

VII A General Model for the Supply of Medical Care

Now that we understand how physicians and hospitals make their choices in this market we can describe a general model of the supply side in the market of medical care. In particular, we focus on the reciprocal effects of the maximizing behavior of hospital administrators and physicians in this market.

In this model, we assume complete and perfect information, so that each hospital administrator knows $C_i\{QN^*, QL^*\}$, for every possible prototype physician i , and every possible set of alternative hospitals, operating at levels of outputs characterized by $\{QN^*, QL^*\}$. The same

information is shared by the set of N physicians active in the market. Note that this has nothing to do with the fact that consumers have imperfect information with respect to prices and quality of care.

a. Quantities, Qualities and Prices

Any hospital j that wants to enter the market at (Q_N^*, Q_L^*) must make sure that there exists at least one physician i for which: $(Q_N^*, Q_L^*) \in C_i \cup \{(Q_N^*, Q_L^*)\}$. If for no physician $i \in N$, is $(Q_N^*, Q_L^*) \in C_i \cup \{(Q_N^*, Q_L^*)\}$, then (Q_N^*, Q_L^*) is not a feasible entry point for hospital j .

Thus, the general equilibrium in the supply of physicians determines both the EE and the FF curves that every hospital administrator faces. In a unidimensional profit maximizing, world this means that every hospital must choose a (Q_N^*, Q_L^*) to allow at least some physicians to realize rents comparable in magnitude to the rents realized by their colleagues, elsewhere in the country. We thus replace the "capture" model with a model where hospitals must choose their output levels to allow physicians to realize rents comparable to existing rents in the market at the time.

In addition, every new hospital entrant affects the entire supply of physicians. Recall that every new hospital forces all incumbent hospitals to change their respective (Q_N^*, Q_L^*) 's. Thus, a new entrant not only attracts physicians from incumbent hospitals, but also alters the entire distribution of physicians across hospitals.

On the other hand, as we discussed in section IV, each physician entrant l to hospital j will directly affect the rents earned by any physicians i working in hospital j . If no such entry occurs at hospital k , $k \neq j$, then any physician i working for hospital j will have

to reconsider his choice $C_i\{(QN_k^*, QL_k^*), (QN_j^*, QL_j^*)\}$, since $u_i(QN_j^*, QL_j^*)$ before the entry of l to j , differs from $u_i(QN_j^*, QL_j^*)$ after l 's entry. Again, as a result of l 's entry, if it turns out that for no physician $i \in N$, is $(QN_j^*, QL_j^*) \in C_i\{(QN_k^*, QL_k^*), (QN_j^*, QL_j^*)\}$, then hospital j will have to change its quality and quantity to survive in the market, since no hospital can operate without its medical staff.

b. Cost Schedules

As we noted above, for every physician i , $MC_i = MC_i(q_i, w_i, QN_j^*, QL_j^*)$, where q_i is the quantity produced by the physician, w_i is cost vector of the vector of input factors needed for production (physician's labor, investment in medical education, etc.), QN_j^* is the equilibrium quantity of each service that the hospital is determined to provide, and QL_j^* is the equilibrium quality that the hospital charges for this service. Thus, every new hospital entrant, or even a shift in the quantity or quality of production of any existing hospital, will affect not only the utility that each physician obtains from practicing in any hospital, but their cost schedules as well.

In the same way, the fact that the hospital must guarantee to its medical staff rents comparable to those observed by physicians elsewhere in the market obviously affects the cost schedules of the hospital. First, and most obvious, rents reserved for the medical staff can be directly treated as an argument in the cost schedules of the hospital. Second, the equilibrium choices of physicians in terms of quantity, quality and location of practice serve as an additional constraint on the quantity and prices of the hospital services. We discuss below how physicians' choices affect the hospital's EE and the FF curves. We want to emphasize here that physicians' choices will determine,

along with other constraints, the hospital's choice of the point on the LRAC at which the hospital will operate. Thus, if it is too expensive to attract physicians to Wichita, Kansas, hospitals there may be constrained to operate in the range where the LRAC is decreasing, and be unable to expand operations, not for lack of patients but because the unacceptable price of attracting more physicians.

c. Demand Curves

Because physicians are free to choose between hospitals to maximize their own objective functions, and hospitals depend on their medical staff for patients, doctors have an important bargaining leverage, which forces the hospital's directorate to have doctors' interests as an argument in its "utility function". The patients that a physician can direct to a hospital represent a source of bargaining power that must affect the amount of specialized resources a hospital devotes to the doctor's particular area of medical care.

Nevertheless, an individual physician-firm cannot expect to capture all of the gains from the business it brings a hospital in general. To understand why, return to Figure 3, for hospital j . If a doctor "brings" more patients to hospital j , this causes the hospital's demand curves to shift outward, thus shifting the EE curve outward (say, from $E'E'$ to EE in Figure 3 as the demand curves shift from D' to D).²⁰ In general, when the bargaining game is at an equilibrium, the apparent preferences of the hospital will balance between competing goals. Thus, if we interpret QL as any activity that augments doctors' private earnings and QN as other activities, then if any doctor moves new patients into hospital j , the EE curve will shift outward, but the extra opportunities afforded to the hospital will be

distributed among competing goals. The "doctor capture" model of Pauly and Redisch is obviously a corner solution to this more general analysis, and one that we would not expect as a general solution to a bargaining game such as this.

A final issue remains in the discussion of interaction between physicians and hospitals, namely whether these two services are complements or substitutes in the production of medical care. To see how this matters, consider the case where physicians and hospital services are complements.²¹ If physician entry produces higher physician prices (as we discussed earlier), then, with complementarity, demand curves confronting hospitals would shift inward, and hence the EE curves would shift (as from EE to E'E' in Figure 3). This would echo back to the physician market, since the hospital opportunities to augment physician earnings (as in the discussion surrounding Figure 7) would diminish. Of course, all of the goals of the hospital would suffer accordingly as the opportunity-possibility frontier FF shifts inward. Obviously, the reverse would hold if doctor and hospital services were substitutes.

VIII. Concluding Remarks

We have established a general model integrating a broad literature in the economics of medical care. We show how equilibrium occurs in markets dominated by not-for-profit firms, and how doctors interact with hospitals in a complex bargaining game. These results generalize previous analyses of NFP hospitals that implicitly or explicitly assumed monopoly power and blocked entry.

This model, with a few standard economic assumptions, corresponds well with the observed changes in the US medical markets in recent

years. First, with better insurance, there has been an explosion of new technology and "quality" (Newhouse, 1988), causing prices to increase dramatically in the hospital sector. Insurance obviously increases the opportunity-possibility set of hospitals by shifting out the demand curves for all levels of quality.

Second, despite an outpouring of new physicians into the markets since the mid-1970s, the prices of physician services have increased while physicians real net earnings have apparently remained flat. These conditions can occur readily in our model when entry leads each firm to produce in the realm where marginal costs fall with output, and these same conditions readily lead to the observed cross-sectional positive relationship between prices and physicians per-capita (Pauly and Satterthwaite 1981).

However, in order to obtain these results, we do not need *ad hoc* models of "target income," "demand inducement," or complicated assumptions about relationships between search costs and the number of physician-firms in a market. All of our results obtain directly from standard economic theory. Occam's Razor prefers such models to those with more elaborate assumptions.

We have not disproved the validity of target income models, the existence of demand inducement, or other previous models of physician behavior. However, we have shown that many of the phenomena commonly thought to be explained **only** by the demand inducement or target income models readily emerge under standard models of monopolistic competition. The only constraint we place on our model is that consumer search is incomplete in physician markets, so that a pure competitive equilibrium does not emerge.

Third, we have provided a formal justification for entry-restrictions in the hospital market (e.g., for CON laws), but also showed the considerable informational requirements necessary to make such controls work. In addition, we have shown how such controls could cause increases in prices and reductions in welfare if poorly operated. We have also provided a model whereby limitations on physician supply **might** be desirable, although we have not investigated market conditions to see if such restrictions would be appropriate. Nevertheless, it no longer appears reasonable to dismiss all calls for entry restrictions as made by puppets of the medical establishment. There do exist conditions under which social welfare, broadly construed, could improve with such entry restrictions.

Finally, we emphasize the importance of consumer search in these models. Search enhances the elasticity of demand confronting all physicians. This, in turn, makes it more likely that equilibrium occurs in the realm where increased supply causes prices to fall, which, in turn, eliminates the desirability of entry restriction. Thus, policy makers interested in controlling the costs of medical care may be better served by investing in activities that facilitate search, rather than in those that block entry by hospitals or physicians.

References

- Arrow, Kenneth J. [1963]: "Uncertainty and the Welfare Economics of Medical Care" American Economic Review, Dec 1963 pp. 941-975.
- Banks, Jeffrey [1985]: "Sophisticated Voting Outcomes and Agenda Control," Social Choice and Welfare, Vol. 1, pp. 295-306.
- Baumol, William J., John C. Panzar and Robert D. Willig [1982]: Contestable Markets and the Theory of Industry Structure, Harcourt Brace Jovanovich, San Diego.
- Becker, Edmund R. and Frank A. Sloan, [1985]: "Hospital Ownership and Performance", Economic Inquiry, Jan. Vol. 23, No. 1, pp. 21-36.
- Boardman, Anthony E. Bryan Dowd, John M. Eisenberg and Sankey Williams, [1983]: "A Model of Physicians' Practice Attribute Determination", Journal of Health Economics, Vol. 2, No. 3.
- Booten, LaVonne and Julia A. Lane, [1985]: "Hospital Market Structure and the Return to Nursing Education", Journal of Human Resources, Vol. 20, No. 2, pp. 184-196.
- Bresnahan, Timothy F. [1981]: "Duopoly Models with Consistent Conjectures." American Economic Review, Dec 1981 pp. 935-945.
- Burstein, Philip and Jerry Cromwell [1984]: "Relative Income and Rates of Return for U.S. Physicians." in: Journal of Health Economics; Vol. 4, pp. 63-78.
- Chamberlin, Edward Hastings [1965; 8TH ed. (1ST edition 1933)]: The Theory of Monopolistic Competition, Harvard University Press, Cambridge.
- Dranove, David [1988]: "Pricing by Non-Profit Institutions: The Case of Hospital Cost Shifting", Journal of Health Economics, Vol. 7, pp. 47-57.
- Evans, Robert [1974]: "Supplier-Induced Demand" in Mark Perlman, ed., The Economics of Health and Medical Care; London, MacMillan, pp. 162-173.
- Feldman, Roger and Bryan Dowd [1985]: "Is there a Competitive Market for Hospital Services?" Journal of Health Economics Vol. 5, pp. 277-292.
- Friedman, Milton and Simon Kuznets, [1945]: "Income from Independent Professional Practice", National Bureau of Economic Research, New York.
- Harris, Jeffrey E. [1977]: "The Internal Organization of Hospitals: Some Economic Implications." The Bell Journal of Economics; Autumn, Vol. 8, No. 2, pp. 467-482.
- _____ [1979]: "Regulation and Internal Control in Hospitals" Bulletin of New York Academy of Medicine; Jan. pp. 88-102.
- Hardin, Russell [1982]: Collective Action; John Hopkins University Press, Baltimore 1982.
- Joskow, Paul L. [1981]: Controlling Hospital Costs: The Role of Government Regulations, MIT Press, Cambridge.
- Kessel, Reuben, [1958]: "Price Discrimination in Medicine", Journal of Law and Economics, Vol. 1, No. 1, pp. 20 - 53.

- Koutsoyiannis, A. [1979]: Modern Microeconomics; St. Martin's Press, New York, 2nd. Edition.
- Manning, Willard G., Joseph P. Newhouse, Naihua Duan et al, [1987]: "Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment", American Economic Review, Vol. 77, No. 3, June, pp. 251-277.
- Marder, William D. and Richard J. Willke, [In Press]: "The Value of Physician Time: Comparisons Across Specialties", in H.E. Frech, Ed.,: Regulating Doctors' Fees: Costs, Competition, and Controls Under Medicare, American Enterprise Institute, Washington.
- McCarthy, Thomas R. "The Competitive Nature of the Primary Care Physicians' Service Market", Journal of Health Economics, Vol. 4, No.2, pp. 93-117.
- McKelvey, Richard D., and P.C. Ordeshook [1984]: "The Influence of Committee Procedures on Outcomes: Some Experimental Evidence", Journal of Politics, Vol 46, pp. 182-205.
- Noether, Monica [1986]: "The Effect of Government Policy Changes on the Supply of Physicians: Expansion of a Competitive Fringe," Journal of Law and Economics, Vol. 29, October, pp. 231-262.
- Newhouse, Joseph P. [1970]: "Toward a Theory of Non-Profit Institutions: An Economic Model of a Hospital" American Economic Review; March, pp. 64-74.
- _____, A.P. Williams, B.W. Bennett, and W.B. Schwartz [1982]: "Does the Geographical Distribution of Physician Reflect Market Failure?" The Bell Journal of Economics, Vol. 13, No. 2, pp. 493-505.
- Olsen, M. Donna et al.: "Medical Care as a Commodity: An Exploration of the Shopping Behavior of Patients" Journal of Community Health; Vol. 2. No. 2 Winter 1976. pp. 85-91.
- Olson, Mancur, Jr. [1971]: The Logic of Collective Action; Harvard University Press, Cambridge Mass., 2nd. Edition, 1971.
- Pauly, Mark V. [1980]: Doctors and Their Workshops, University of Chicago Press, Chicago.
- Pauly, Mark, V. and Michael Redisch [1973]: "The Not For Profit Hospital as a Physicians' Cooperative" American Economic Review, Vol 63, No. 1, march, pp. 87-99.
- Pauly, Mark, and Mark A. Satterthwaite [1981]: "The Pricing of Primary Care Physicians' Services: a Test of the Role of Consumer Information" The Bell Journal of Economics, Autumn, Vol 12, No. 2, pp. 488-506.
- Phelps, Charles E. and Joseph P. Newhouse [1972]: "Coinsurance and the Demand for Physician Services," Social Security Bulletin, June.
- Phelps, Charles E. and Joseph P. Newhouse [1974]: "Coinsurance, the Price of Time, and the Demand for Medical Services", Review of Economics and Statistics, August, Vol. 56, No. 3.
- Rose-Ackerman, Susan, [1987]: "Ideals Versus Dollars: Donors, Charity Managers, and Government Grants, Journal of Political Economy, Vol. 95, No. 4, pp. 810-823.

- Rose-Ackerman, Susan, Ed. [1986]: "The Economics of Nonprofit Institutions, Oxford University Press, New York.
- Sadanand, Asha, and Louis Wilde [1982]: "A Generalized Model of Pricing for Homogeneous Goods under Imperfect Information", Review of Economic Studies; vol. 49, pp.229-240.
- Salkever, David S. and Thomas W. Bice, [1976]: "Impact of State Certificate of Need Laws on Health Care Costs and Utilization", Milbank Memorial Fund Quarterly: Health and Society, Vol 54, Spring, pp. 185-214.
- Schotter, Andrew [1981]: The Economic Theory of Social Institutions, Cambridge University Press, Cambridge.
- Schwartz, Alan, and Louis L. Wilde [1982]: "Competitive Equilibria in Markets for Heterogeneous Goods Under Imperfect Information: a Theoretical Analysis with Policy Implications": Bell Journal of Economics; Spring, Vol. 13, No. 1; pp. 181-193.
- _____, [1982]: "Imperfect Information, Monopolistic Competition and Public Policy": American Economic Association, Papers and Proceedings; May, Vol. 72, No. 2; pp. 18-23.
- Shepsle, Kenneth [1979]: "Institutional Arrangements and Equilibrium in Multi dimensional Voting Models", American Journal of Political Science; Vol. 23, pp. 27-59.
- Shortell, S.M., Morrison E.M., Hughes S.L. [1986]: "The Effects of Hospital Ownership on Nontraditional Services" Health Affairs Winter, Vol. 5, pp. 97-111.
- Sloan, Frank A. and Bruce Steinwald [1980]: "Effects of Regulation on hospital cost and Input Use" in: Journal of Law and Economics; Vol. 23, No. 1, pp. 81-109.
- Sloan, Frank A. [1983]: "Rate Regulation as a Strategy for Hospital Cost Control: Evidence From the Last Decade"; Milbank Memorial Fund Quarterly; Vol. 61, No. 2, pp. 194-214.
- Stano, Miron [1987]: "Analysis of Physician Inducement Controversy" Journal of Health Economics; Vol. 6, pp. 227-238.
- Weisbrod, Burton A. [1988]: The Nonprofit Economy; Harvard University Press, Cambridge.
- Wilde, Louis L and Alan Schwartz [1979]: "Equilibrium Comparison Shopping," Review of Economic Studies; Vol. 46, No. 3; pp. 543-553.

Endnotes

- ¹ For an excellent assessment of the characteristics of the NFP sector in the U.S., as well as its financing, see Weisbrod [1988].
- ² In addition, the Internal Revenue Service exempts NFP organizations from paying federal income tax [through Sec. 501.c.3 of the IRS code], and most state and local governments join with a parallel exemption from paying state income and sales taxes, local property taxes, etc.
- ³ Dranove (1988) uses a similar model with multiple classes of buyers (each with a different demand elasticity) to provide a theoretical model that explains "cost shifting".
- ⁴ Note that with sufficiently strong preferences for QN, the NFP hospital would be willing to operate in the realm where demand is inelastic, and still would charge a positive price. This could occur if $U_{QN}/\lambda > MC$ and $|\eta| < 1$. No for-profit monopolist would ever willingly operate in this range of a demand curve, because MR is negative there.
- ⁵ For our purposes here, consider this as a long run average cost curve. As we later show, the implications of short run disequilibrium do not alter much the implications of this model, and the nature of the "equilibrium" in the model is more consistent with the LRAC interpretation.
- ⁶ If hospitals produce vectors of qualities, then the cost of production is a blend of the costs of these vectors. We can meaningfully think about a single AC curve so long as the mix of qualities remain constant along the overall expansion path, i.e., the "ray-average costs" of Baumol, Panzar and Willig [1982] in their discussion of multiproduct firms.
- ⁷ The usual mechanism for a local government to issue tax exempt bonds, and then use the proceeds to lend to the NFP hospital in a direct passthrough of the lower interest rate.
- ⁸ Note also that, particularly in this case, nothing ensures that the EE curve ever bends backward. Even in this case a unique equilibrium exists if we assume (reasonably, we believe) that there exists some maximum quality QL_{max} , such that for $QL > QL_{max}$, demand curves never intersect AC curves. This terminates the EE curve at a specific point. The comparable FF curve also slopes upward, and terminates at the same quality. A utility-maximizing decision maker will produce at this point, a corner-solution "tangency" to the family of indifference curves of the decision maker.
- ⁹ The ways in which the rules established bed limits varied, but commonly, they calculated an approved level of beds from some formula that included information about each region's age and sex distribution. For description of these laws, see Joskow, 1981.

¹⁰ One study by Shortell et al. (1986) finds NFP hospitals more likely to provide "non traditional" services such as geriatric care, home health services, and health promotion.

¹¹ Recall that by "quality" we mean any dimension of the hospital's performance not directly measurable as "quantity, in our "two goods" case. Thus, "quality" need not measure "medical quality" in a sense that physicians would ordinarily use this term. It represents any component of the hospital's utility function. For example, it might represent transfers to physicians, higher wages to any subset of employees, etc.

¹² In a recent paper, Stano [1987] has noted that: "Except for the assumption of higher individual demand elasticities, the literature on physician markets has not developed a complete monopolistically competitive model."

¹³ In our model, humans may be conceptualized as sets $\{HS\}$ [health status], s.t. a defect in each $hs_i \in \{HS\}$, is an $hp_i \in \{HP\}$ [health perturbation], with a corresponding $hr_i \in \{HR\}$, [health repair] s.t. for any $\{HS\} = \{hs_1, \dots, hs_{i-1}, hs_i, hs_{i+1}, \dots, hs_n\}$ turned into an $\{HS'\} = \{hs_1, \dots, hs_{i-1}, hp_i, hs_{i+1}, \dots, hs_n\}$, there exist an $hr_i \in \{HR\}$, s.t. $\{HS'\} + hr_i = \{HS\}$. It is unlikely that two physicians will administer exactly the same hr_i , when facing a similar hp_i . This leaves us with a market of essentially similar but not identical goods. This urges us to treat physicians as involved in "monopolistic competition" over the provision of similar 'hrs'.

¹⁴ The idea of barriers to entry in medical care can be traced to Friedland and Kuznets (1945), Kessel (1958) and Arrow (1963). All of these authors emphasize the potential for licensure to limit entry to the profession as a whole, yet none explain how monopoly power could persist at the final product level. Professional licensure restricts entry into an input market (physician labor). There exist some evidence that entry has been restricted into the medical profession. The return to medical education considerably exceeds competitive returns (Mardor and Wilke, 1989), even after adjusting for hours of work (Mitchell and Cromwell, 1984). Further, the returns to specialization persistently seem higher for surgical specialties, where the opportunities for substitution almost certainly fall to or near zero for at least some of the activities of the physician-firm.

Legal constraints, if not technical constraints, prevent RNs or PhDs in economics from performing neurosurgery. Medical malpractice laws probably also inhibit some forms of substitution that might be technically feasible.

The substantial increase in demand for medical services in the 1960s and 70's following the passage of Medicare legislation induced a significant influx of foreign-trained physician, to the point where over half of newly licensed physicians came from foreign medical schools during some years in the 1970s (Noether, 1986).

¹⁵ Sadanand and Wilde [1982] characterize the set of monopolistic competitive equilibria using levels of price search as a parameter. They show that above a certain amount of search in the market, the monopolistic equilibrium "collapses" to the competitive market equilibrium. Below this level of search, the set of equilibria is a continuum bounded by the monopolistic equilibrium. When search is reduced, the set of equilibria includes more and more points from the neighborhood of the monopolistic equilibrium and less and less points from the neighborhood of the competitive solution. In section Va. below, we explain why our analysis is consistent with any of these possible equilibria, save the purely competitive market equilibrium. See also Schwartz and Wilde, 1982.

¹⁶ For a rigorous discussion of this and related equilibrium concepts, see Koutsoyiannis [1979] pp. 202-233.

¹⁷ In general, we see no reason why normal search strategies by consumers would cause search costs (and hence optimal search) to vary with the number of doctors in the community, in contrast with the assumed positive relationship in Pauly and Satterthwaite.

¹⁸ See again endnote 13. In the health care industry, we can expect that search will occur much more frequently for routine and ambulatory care than for specialized and infrequently occurring events such as surgical care. This may help explain why rates of return to some specialty training (such as family medicine and pediatrics) is very small (Mardor and Willke, 1989; Burstein and Cromwell, 1986). Rates of return in specialties not confronting much search could persist at higher levels, particularly if entry were constrained at the specialty level, e.g., through limitations on specialty training positions. In addition, we discuss below ways in which the hospital can also create monopoly returns to some specialties through coordinating the use of a common and complementary factor of production (e.g., surgical suites).

¹⁹ For a more elaborate discussion on the issue of how the 'collective action' problem shades doubts on the validity of most oligopolistic models see Schotter 1981, pp. 41-42. On the particular difficulties that the collective action problem raises with respect to Chamberlin's model of 'monopolistic competition' see Hardin 1982, p. 126-131.

²⁰ This does not necessarily represent "demand inducement" because these shifts represent changes in patients' choices across hospitals, akin to the effects of consumers' search.

²¹ Manning et al. 1986 report that on the demand side, complementarity appears to exist broadly. That is, when the price of ambulatory care increased, the demand for hospital care decreased in the RAND Health Insurance Experiment.

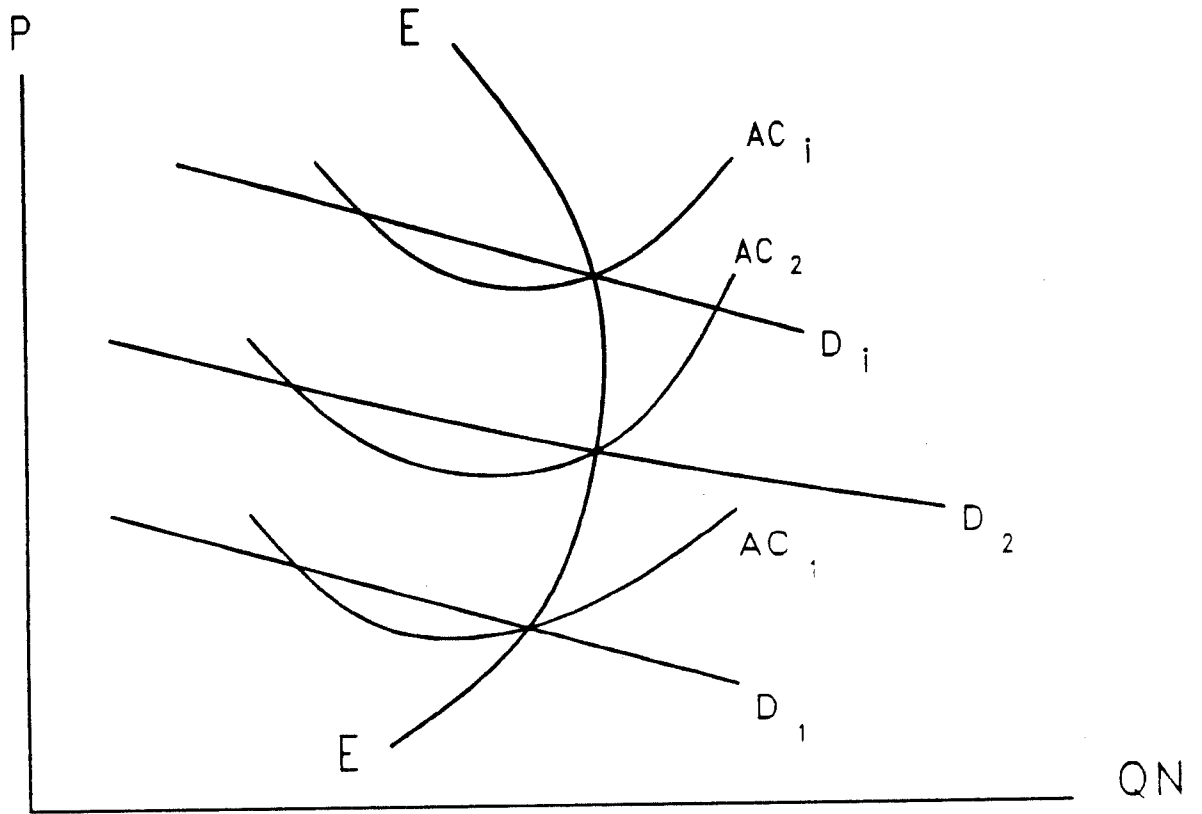


FIGURE 1

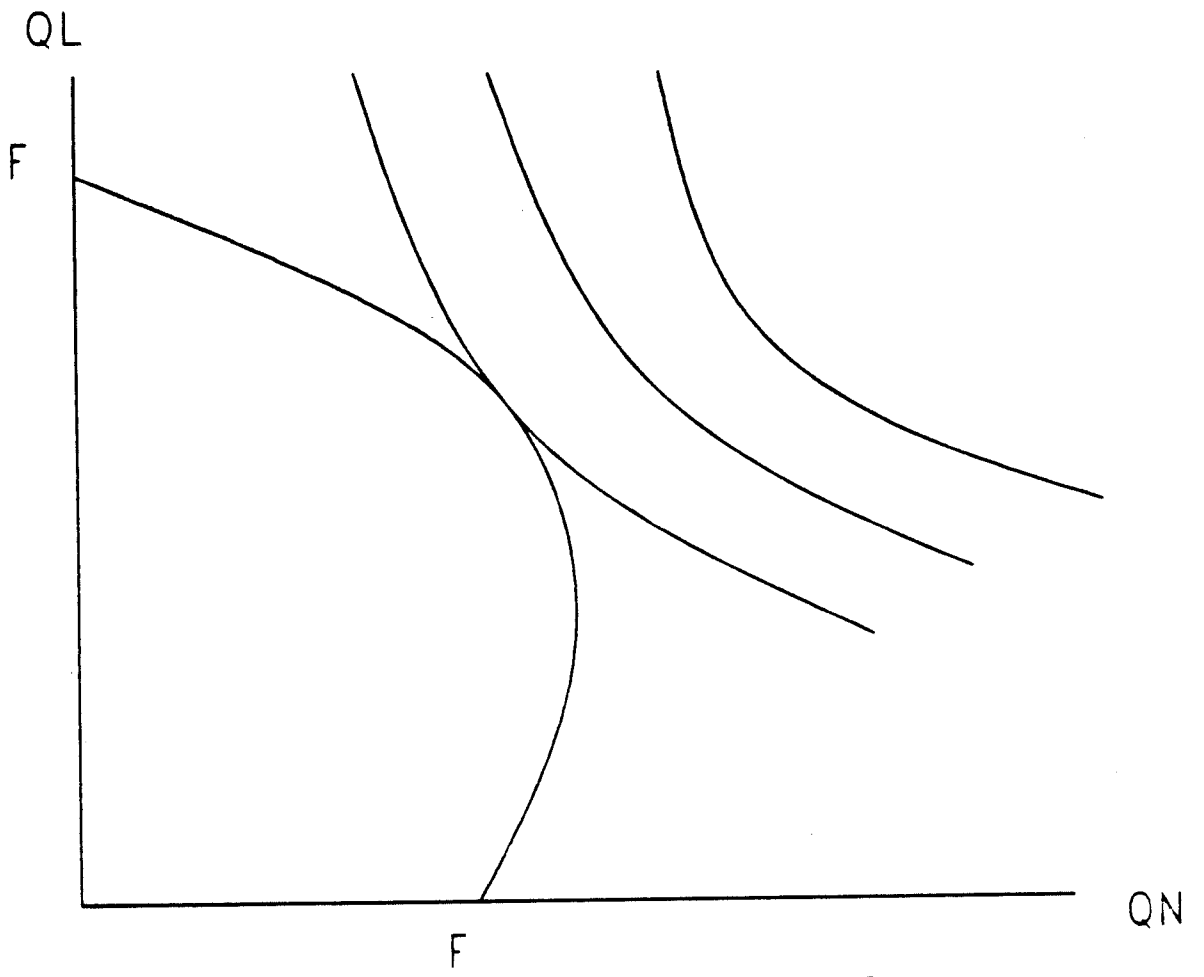


FIGURE 2

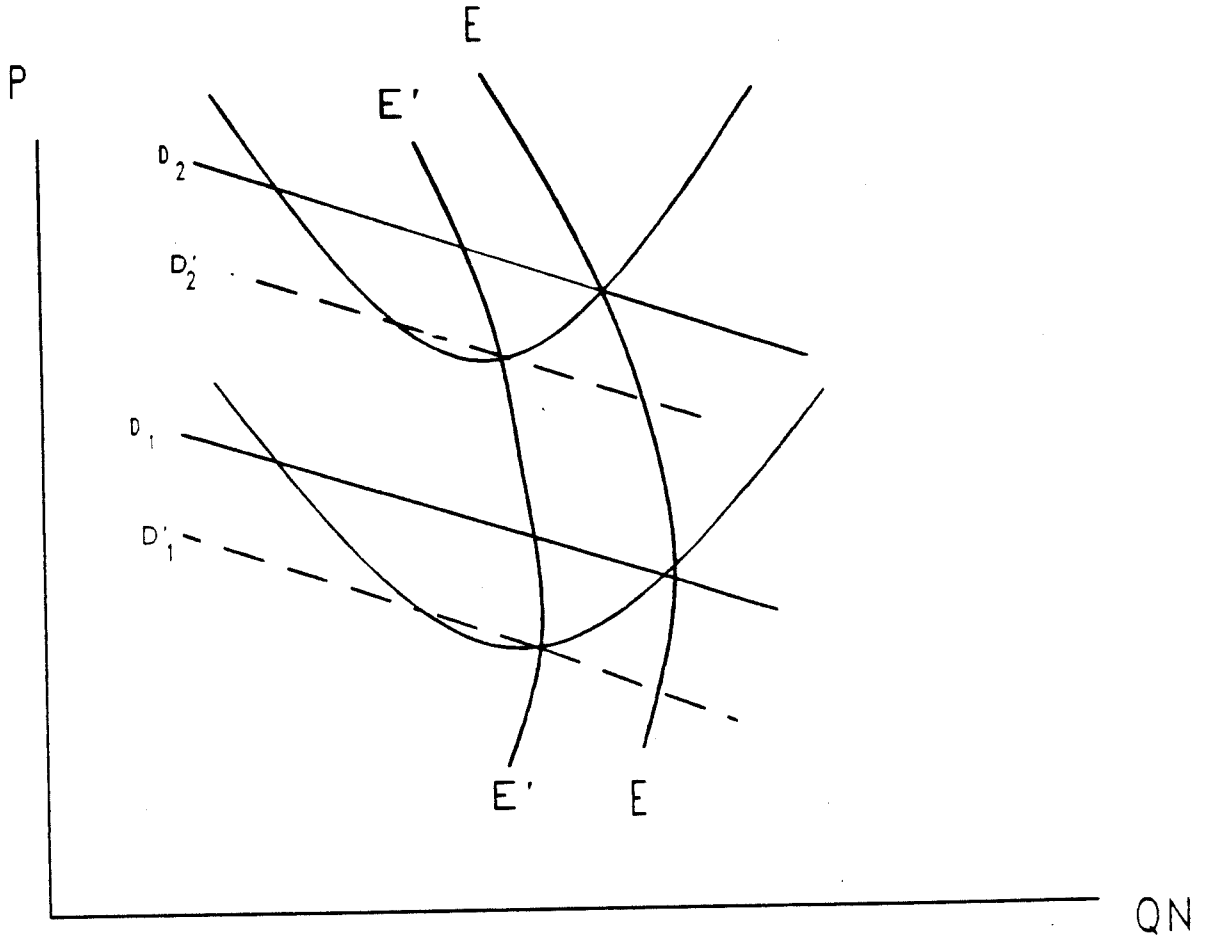


FIGURE 3

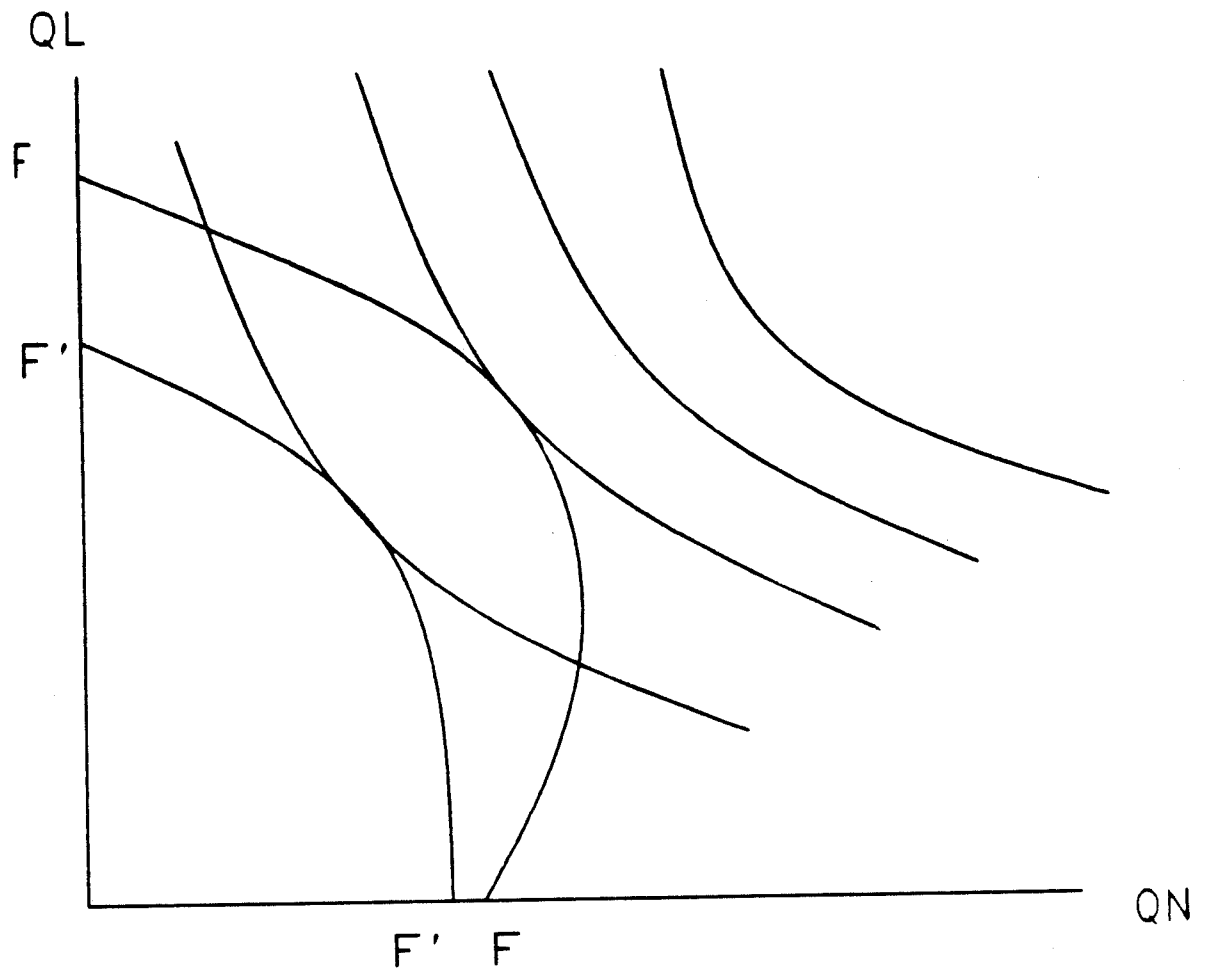


FIGURE 4a

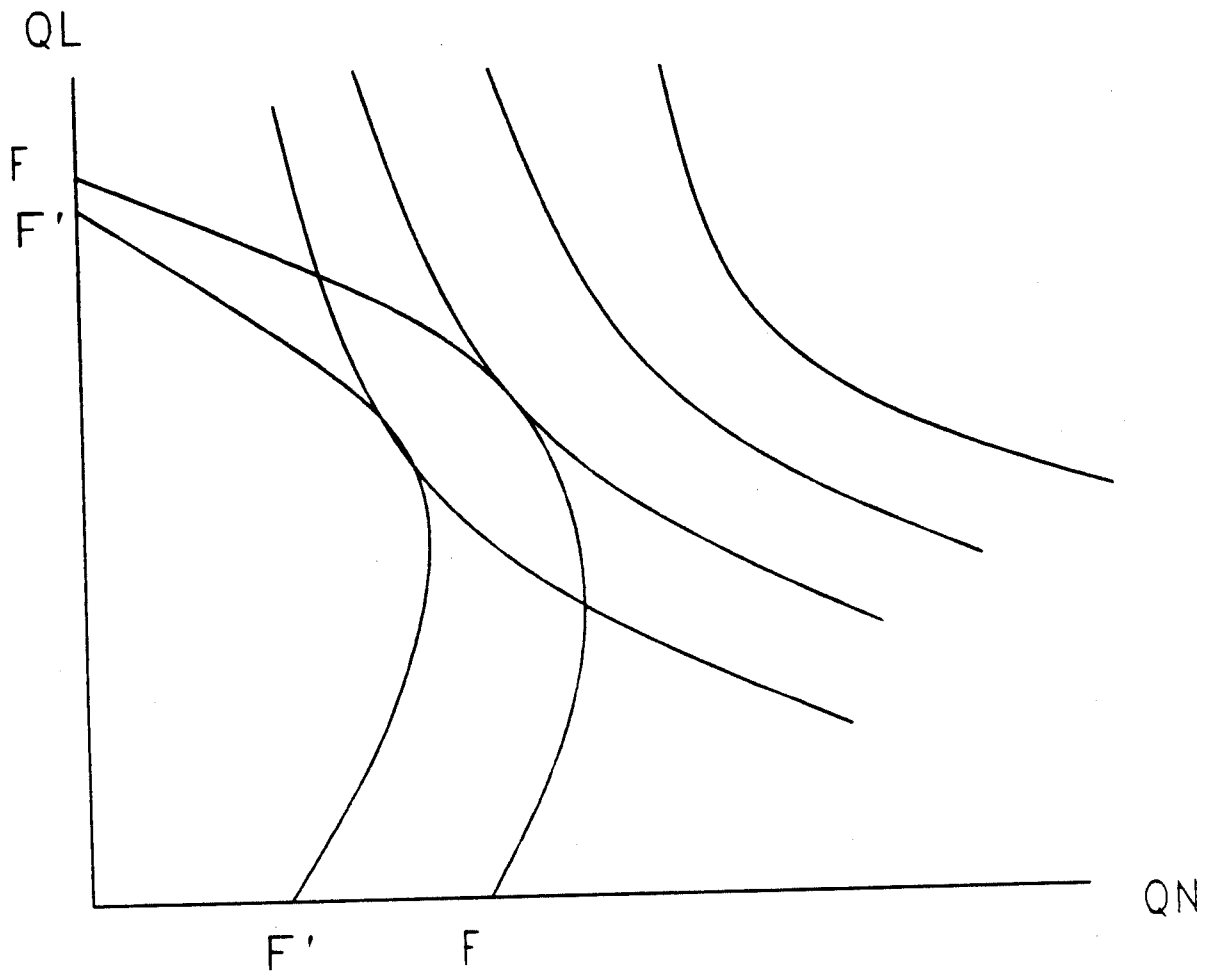


FIGURE 4b

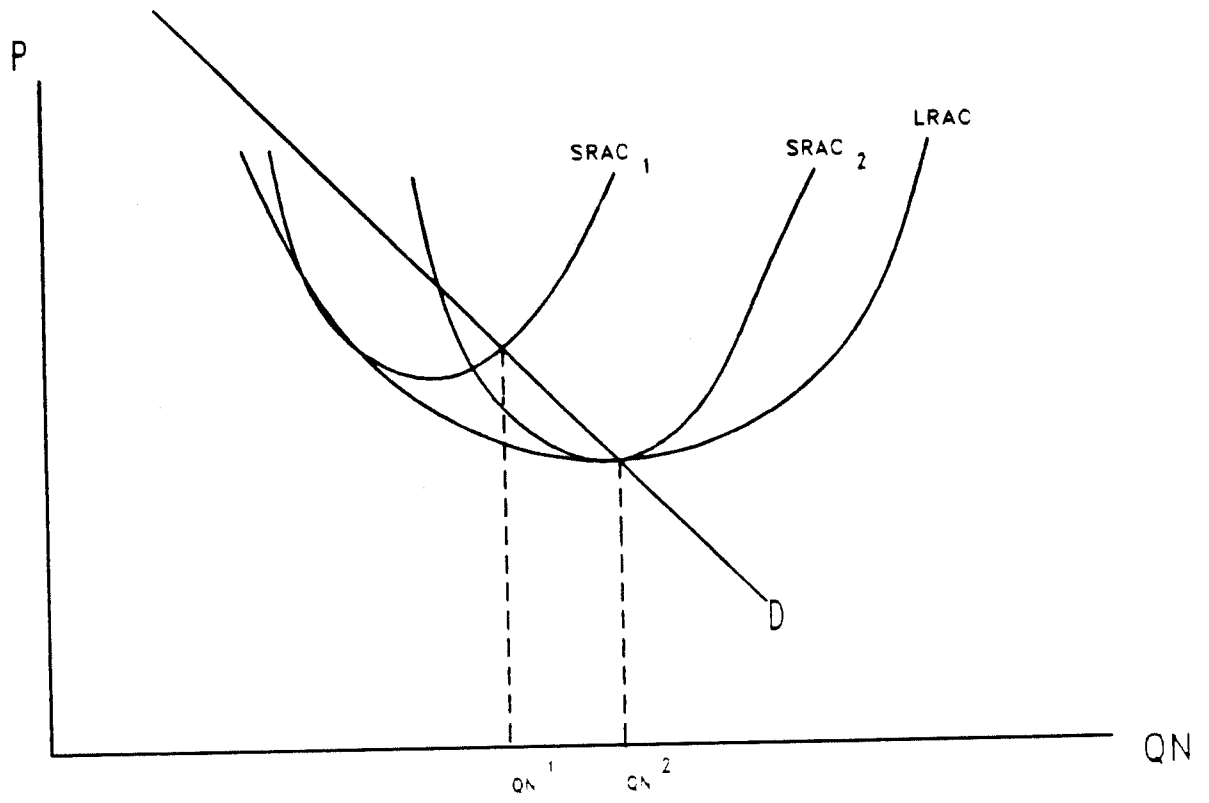


FIGURE 5

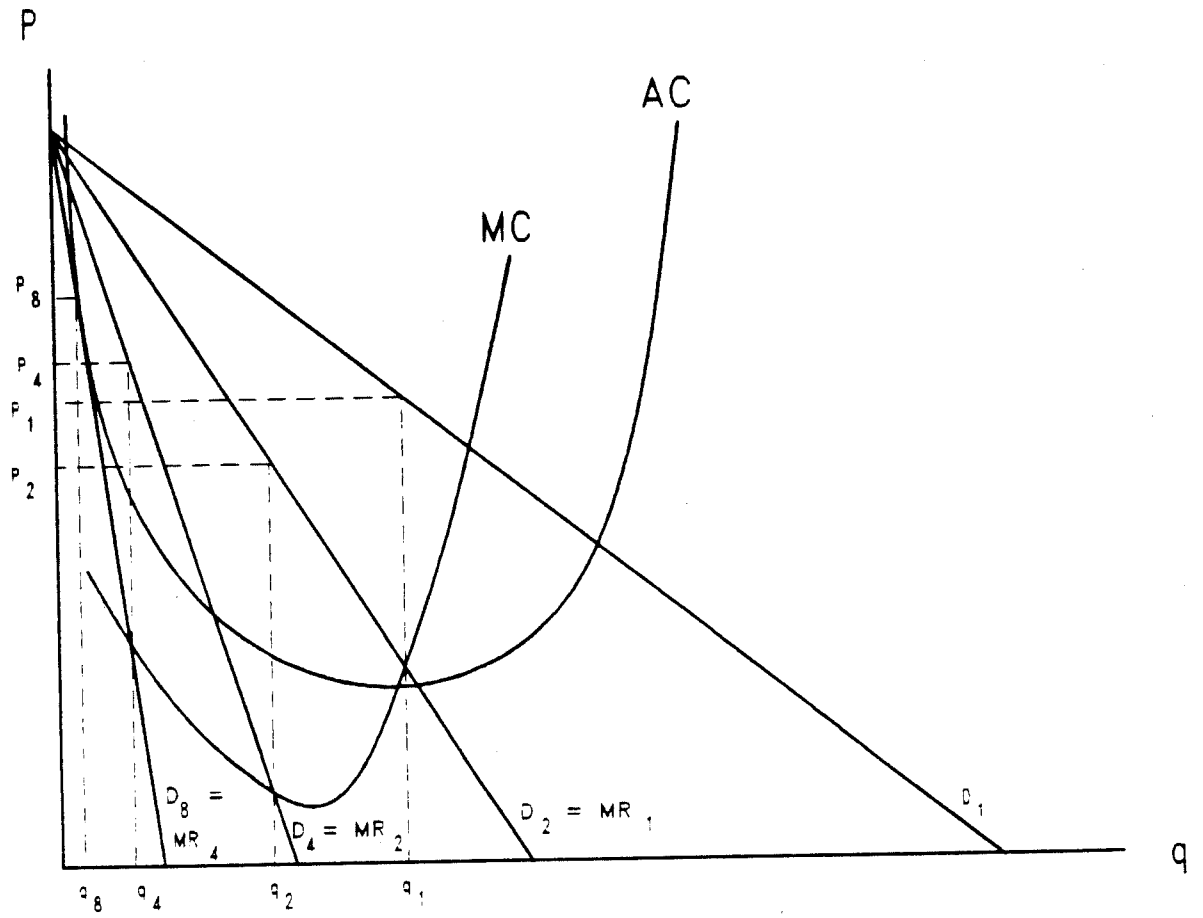


FIGURE 6a

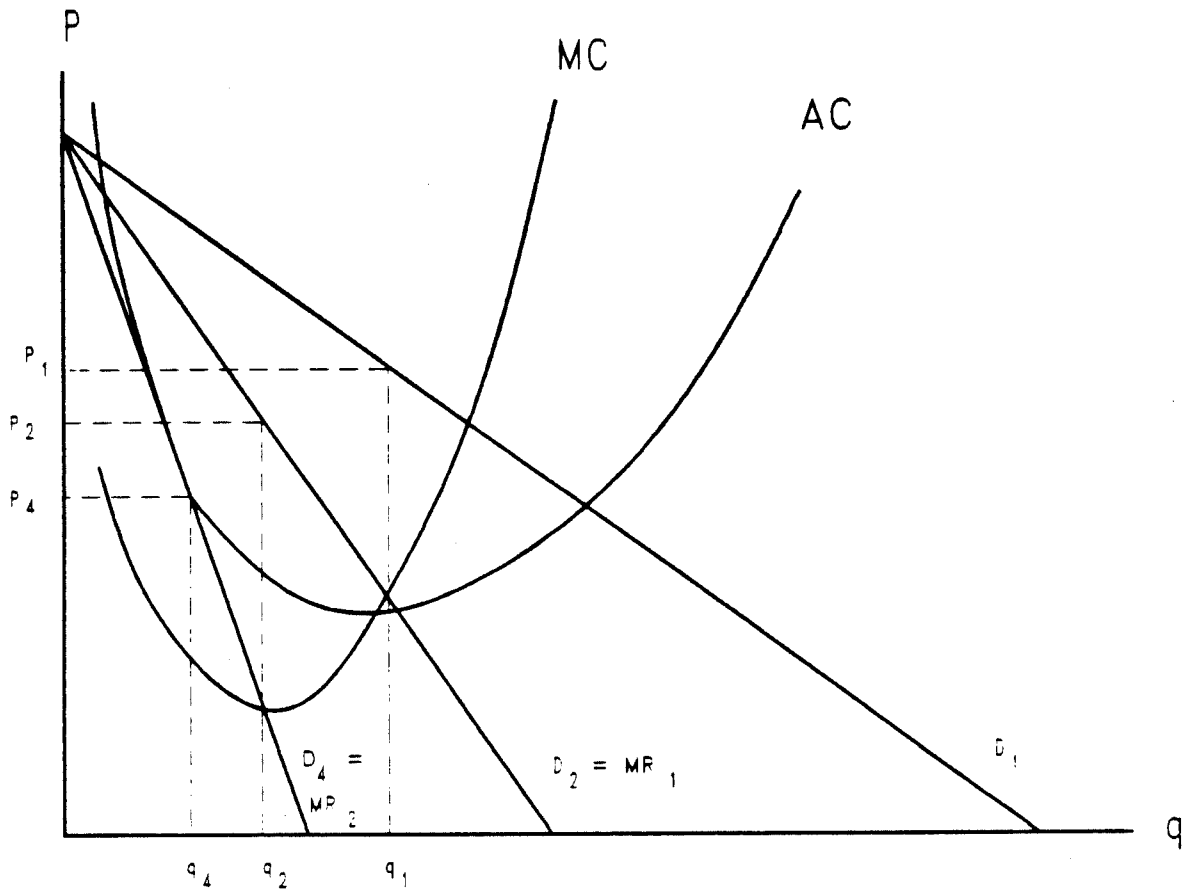


FIGURE 6b

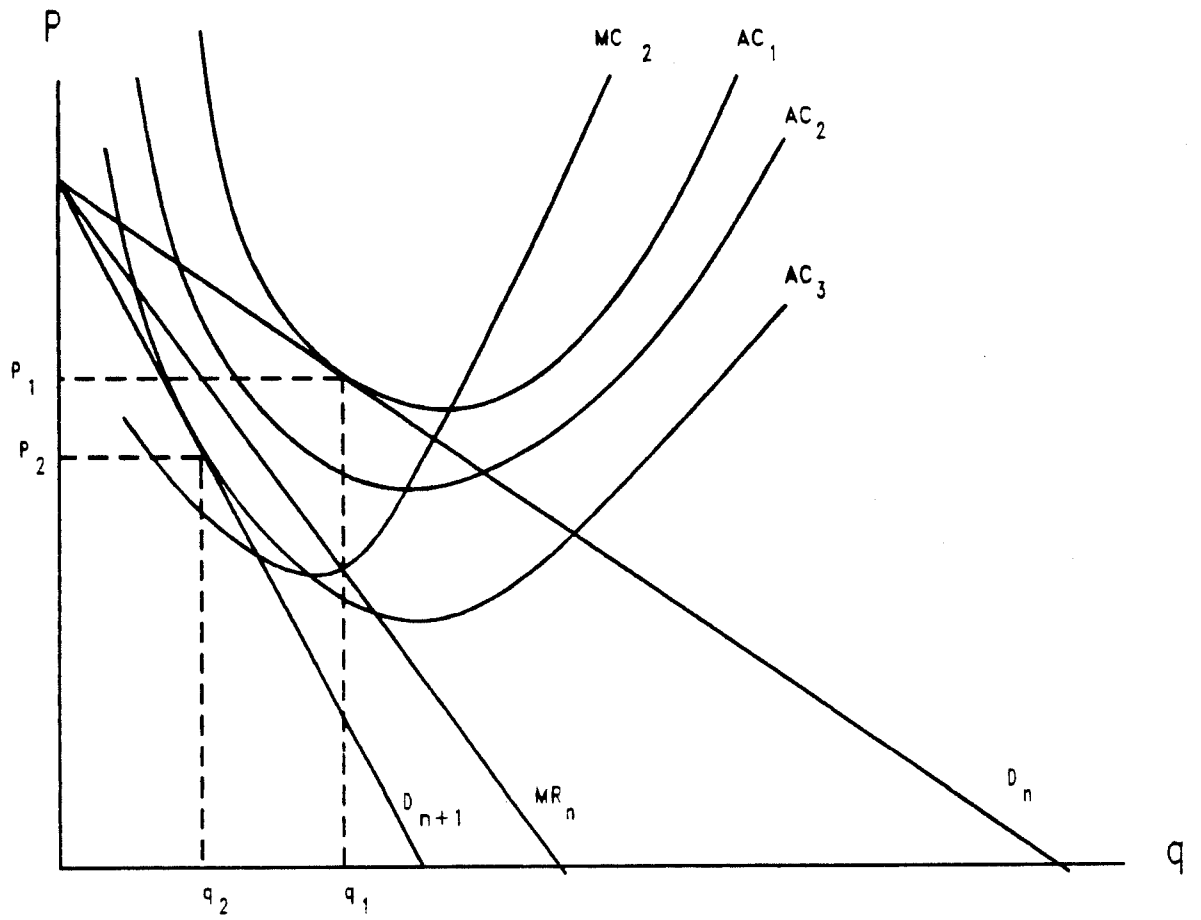


FIGURE 7