INTERNAL PRICING AND COST ALLOCATION IN A MODEL OF MULTIPRODUCT COMPETITION WITH FINITE CAPACITY INCREMENTS

Uday Karmarkar
Richard Pitbladdo

William E. Simon Graduate School of Business Administration
University of Rochester
Rochester, New York 14627

QM Working Paper # 90-06

Research support from the Xerox Faculty Fund and the Center for Manufacturing and Operations Management is gratefully acknowledged. Special thanks are due to Jerold Zimmerman for helpful comments on earlier drafts.
Internal Pricing and Cost Allocation in a Model of Multiproduct Competition with Finite Capacity Increments

Uday Karmarkar
Richard Pitbladdo
William E. Simon Graduate School of Business Administration
University of Rochester
Rochester, New York 14627

ABSTRACT

Internal prices are used in practice to allocate central resources to a firms' profit centers. The fixed costs of capacity acquisitions are often included in these prices. We examine the interaction between capacity acquisition and competition when capacity is available in fixed increments. We find predictably that if the increments are small, unit capacity cost is a good approximation for the internal price, and if the increments are large, the internal price is zero. However, the relationship between the internal price and the capacity increment for intermediate cases is quite irregular, to the extent that it is not possible to approximate the internal price with accounting data. The analysis also suggests that full cost allocation overcharges for the opportunity cost of capacity. Furthermore, the right internal price does not act either as a way of recovering fixed costs or as a proxy for externalities such as congestion costs. The conclusions are not materially altered in the case where variable costs increase at the margin, and where these costs rather than hard capacity constraints are the reason to restrict output.
Introduction

Many firms which produce a variety of products appear to be reorganizing to be "product focussed" rather than "process focussed". This is presumably done to orient management toward the special aspects of the marketplace for each product. A decentralization into product-based profit centers reduces the burden of simultaneous consideration of all products. Yet such consideration is typically inescapable, if for no other reason than that the various products share some resources. Presumably, this is why the products were produced together in the first place. Examples of such capacitiated resources include melting equipment for glass or metal, wave solder and automatic insertion equipment for printed circuit boards, and injection molding equipment in plastics manufacturing.

In this context, the various products need to be considered simultaneously at two junctures, the first being the decision of whether and to what extent to acquire common resources. Subsequently, to the extent that the different products compete for limited capacity of the common resources, the merits of the demands of each product for capacity must be reconciled in some way. The first question of resource acquisition has been considered in Karmarkar and Pitbladdo(1990) in the special context of common resources without capacity limitations. The emphasis there is on the effectiveness of common accounting procedures for estimating product line profitability in supporting product line decisions.

The broader motivation behind this study and its predecessor bears some explanation. Changing conditions in manufacturing technology are resulting in a changing picture of the cost structure of manufacturing. To take an extreme example, the appearance of highly automated, integrated manufacturing systems, creates substantially higher fixed to variable cost ratios in comparison with conventional manufacturing techniques (Lederer and Singhal, 1988). Such systems are also less amenable to small capacity adjustments at the margin. In other words, capacity constraints are a significant issue, so that resource allocation and quantity decisions gain in importance. The flexibility of the newer technologies also intensifies the multi-product character of the problem, since a much wider range of products can be produced on the same system. Jaikumar
(19xx) has characterized these issues in terms of determining the portfolio of products that will be produced on these systems. Insofar as cost allocation and internal pricing of resources are related to these decisions, the importance of these practices is growing and errors can become more critical.

Here we shall examine the ability of accounting data to support the decisions regarding the production quantities of each chosen product in the presence of limited capacity. The model has one period of production, and therefore does not distinguish between initial and recurring fixed costs. However, suitably interpreted, it applies to either, given a stationary environment. The question of acquisition remains an integral part of the analysis, but we defer to Karmarkar and Pitbladdo (1990) on the issue of supporting such decisions with accounting data. The next section briefly reviews the results of that paper. In the following section, the main findings of this paper are summarized. The subsequent section presents a general analysis of the question, omitting technical derivations. These are relegated to an appendix.

There is a very large research literature spreading across accounting, economics and management science that addresses the related issues of resource pricing, transfer pricing, resource allocation, and allocation of common fixed costs. The conceptual paradigms and methods used to pose and analyze the issue also vary widely, ranging from traditional economic analysis of competitive situations, and game theory to mathematical programming models, and procedural accounting heuristics. By way of providing the reader with a brief introduction, we summarize a small part of the recent literature. In the area of practice, Kovac and Troy (1989) describe Bellcore's approach to transfer pricing, which is essentially full cost allocation based on the relative rates of usage. Scarpo (1984) describes transfer pricing practice in auto dealerships. Bost (1986) discusses the CASB standards and framework for full allocation practice with reference to two defense industry cases involving McDonnell Corp. and Lockheed Corp.

Management science models of pricing and allocation typically involve more serious efforts at modelling production functions. Balachandran, Kashi and Srinidhi (1990) extend the Miller and Buckman (1987) analysis of allocating the fixed cost of common capacity in a queueing


Economic models are closest to the approach followed in this paper. Dorward (1986) studies full cost versus marginal pricing with imperfect information in an oligopolistic environment. He notes that 63% to 83% of US and UK firms use ex ante overhead allocation to develop costs for pricing. He develops full cost allocation heuristics appropriate for collusive and non-collusive oligopolies. Lee (1984) surveys the long standing debate between marginal and full-cost pricing, which underlies the topic of this paper and its predecessor. He asserts that the demise of full-cost pricing is more the result of doctrinaire opposition than fundamental analysis.

The contribution of this paper from a methodological viewpoint, is in the development of a tractable oligopoly model of multi-product firms with fixed costs of capacity acquisition and constrained production capacity. This framework permits the analysis of market prices, internal prices for common constrained resources and their relationship to common fixed costs.

The Uncapacitated Case

In certain circumstances, there are common fixed costs associated with the production of multiple products without associated limitations on capacity. Examples fixed transactions costs of acquiring and maintaining technology and resources, costs of research and development, and costs
of resources such as land and buildings that in many cases do not constrain production volumes. Karmarkar and Pitbladdo analyze a Cournot-Nash model of multiproduct firms competing in these circumstances. The situation is modeled as a two stage process with firms making an entry decision at which the fixed costs are incurred, and then making decisions on production quantities so as to maximize profits, contingent on the number of firms in the marketplace.

In the second stage of the game, although there are no capacity limitations, firms have an incentive to limit production so as to raise prices. As in the extreme case of a monopoly, the profit maximizing price is in general not equal to marginal costs of production. The analysis of the second stage establishes the relationship between the number of firms in the market, and the profitability of the individual firm. Stepping back to the entry decision, firms will incur the fixed costs and enter as long as prospective profits (or rather contribution in excess of variable costs) exceed the fixed entry cost. The effect of the fixed cost is thus to limit entry. To a first approximation, it then happens that the fixed costs of entry are covered by contribution margin.

Superficially, prices appear to be set so as to "cover fixed costs". However, the underlying decisions are profit maximizing and are not reached by setting prices to cover allocated fixed costs plus variable costs. In fact, the analysis of the basic setting reveals no reason to allocate costs to individual products. Karmarkar and Pitbladdo go on to argue that in more complex cases, cost allocation may provide a useful heuristic. They suggest that allocation proportional to contribution is a rule that adequately mimics the "correct" analysis, and has the potential for providing a workable heuristic for complex problems. This direction is pursued further by Pitbladdo (1990).

The analysis of the uncapacitated case does not shed much light on the effect of capacity constraints. In the latter case, the need for allocation of available capacity to individual products creates interactions that do not exist in the former case. It could be that fixed costs as well as capacity limits must be considered. For example, the allocation of capacity suggests the use of prices for capacity as an allocation mechanism; do these prices have the effect of allocating fixed costs? To what extent are these prices related to fixed costs? It is to these questions that this paper is addressed.
An Overview of the Main Results

One method used to balance the competing demands for common capacity is to establish an internal service price for the common resource. Each profit center is charged by the central organization in proportion to its use of capacity (which for simplicity we take as being one-dimensional). With this approach, the issue of coordinating the production plans of the various product lines is reduced to the determination of the appropriate internal price. In theory, the price one profit center pays on the margin for capacity should be equal to the foregone opportunity of profit from the next most worthy use of the capacity. Yet to literally apply such a criterion, the central administration must have detailed knowledge of the marketplace for each product. As a result, the central organization would have to take back much of the burden it sought to decentralize in the first place. If the appropriate internal price can be approximated by readily available accounting data, this would considerably ease the burden of central administration and thus enable the firm to capture the full benefits of its product oriented organization plan. Zimmerman (1979) recognizes this consideration and suggests it as a potential explanation for why firms allocate overhead costs (read acquisition costs of common resources) to products.

Indeed, if the efficient scale of the common resource is very small, then capacity can be considered to be available in any quantity at constant unit costs. Differing levels of demand for capacity by any product can be met at the margin by simply increasing or decreasing the level of purchased capacity. In this case it is clear that the internal price should be equal to the unit cost of capacity, which is tantamount to fully allocating capacity costs to the divisions in proportion to capacity utilization. But it should also be clear that if the efficient scale is so small, there is no reason that the resource be common to the various products. Each profit center could separately acquire its own capacity without affecting total capacity acquisition costs (we are implicitly assuming a stable deterministic environment without, for example, statistical economies of scale from sharing resources).

Nonetheless, an important lesson is gained from the small-scale case. Once firm-specific capacity is in place, its cost is sunk. In principle, therefore, the cost of capacity should play no
direct role in the setting of the internal price for operational decisions. The firm's level of capacity, however, is quite important; the higher the capacity, the lower the optimal internal price. But absent uncertainty, if the capacity is selected optimally, the optimal internal price from an operational perspective happens to equal the unit capacity costs (We study a one-period model. For the analogy to hold in a multiperiod setting, the environment would have to be stationary. Then unit costs for durable capacity would be imputed interest costs divided by capacity.).

What if the efficient scale for the resource is not small, so that there is potential reason for it to be shared? For the sake of definiteness, suppose that capacity is available in fixed increments of nontrivial size. If the capacity increment were sufficiently large relative to the size of the marketplace, excess capacity will result. That is, firms would find it in their best interest to produce quantities which do not exhaust available capacity. There would be no conflict among products for capacity, and the optimal internal price, assuming away wear and tear, etc., would be zero. Thus as the capacity increment increases, holding unit capacity cost constant, the optimal internal price starts in the vicinity of unit capacity costs and in some cases eventually declines to zero (it may be that unit capacity costs are high enough so that the increase in capacity increment causes all firms to drop out before the internal price reaches zero). At both extremes, we see that the internal price is easily obtained from accounting data. By carefully studying the intermediate cases, we attempt to find out whether this is true in general.

To study the intermediate cases, we employ a model of interfirm competition. Following Karmarkar and Pitbladdo (1990), a natural choice is a capacity acquisition stage followed by Cournot competition (which will be explained in detail later). This model enables us to predict the distribution of capacity in the industry, the market prices and quantities for all products, and the shadow price of capacity for each firm, which is also the optimal internal price. Assuming independent, linear, downward sloping demand curves in the product market, we find that, for a given number of firms, the internal price falls as the capacity increment increases. This is because the extra capacity drives down product prices and marginal revenue as more product must be absorbed by the marketplace. As operating profits fall far enough so that recovery of fixed cost is
endangered, the market will support one less firm, aggregate capacity will discretely fall, and the optimal internal price will discretely increase.

Under further assumptions, the lower bound attained at these shakeout points turns out to be linear and decreasing in the size of the capacity increment. It is computed primarily by using accounting data, but also requires some information regarding the general level of price sensitivity of the products' demand curves. However, the size of the increases in the optimal internal price at the shakeout points become ever larger as the capacity increment increases. This is simply because the removal of one firm becomes a bigger event when the firm would have had greater capacity. Indeed, at each shakeout point, the internal price returns to the vicinity of the unit capacity costs (until excess capacity appears). Thus the above-mentioned, well-behaved, relatively easily calculated lower bound for the internal price is not at all a good approximation in general.

It is edifying to contrast the results of the analysis of this paper with a more conventional view of the shadow price of capacity derived from an optimization paradigm. First it becomes clear that in a competitive environment, where prices depend on the total quantity brought to the market, the shadow price depends on the total capacity in the marketplace available for production, and not just that of one firm. In other words, if any one firm were to increase its capacity for whatever reason, the shadow price or internal price of capacity in all other firms, would drop. Secondly, the shadow price depends only weakly on the fixed costs of acquiring capacity. Consequently, we do not find support for the suggestion made by Zimmerman that cost allocation is a means of rationing scarce capacity. In fact, basing internal prices on full costs almost always overcharges for the opportunity cost of capacity. In other words, the fixed costs of capacity acquisition are not "covered by" or "recovered from" the internal price. Instead, much as in the uncapacitated case, they are related to the total excess contribution across all products, resulting from restricted entry. The exact proportion of the contribution made by each product is irrelevant, and no allocation of the fixed cost to individual products is necessary.

Finally we also consider the situation of marginally increasing variable costs due to phenomena such as congestion costs or other diseconomies of volume production. We find that identifying the
optimal level of production still depends on total capacity in the marketplace, and that the fixed cost of capacity acquisition is not directly related to the marginal variable cost at the equilibrium production level. Again, the conjecture that cost allocation may function as a proxy for congestion externalities, is not supported.

From the viewpoint of practice, the discouraging result is that the optimal internal price is an erratic function of the capacity increment, requires detailed knowledge of the marketplaces for each product, and therefore cannot be reasonably approximated with internal accounting data. The lesson for management is that it appears that internal pricing is not a simple substitute for the direct coordination of production plans across products, whenever common resources are important, and when these resources are acquired in indivisible increments.

The question of why firms allocate fixed costs remains. In Karmarkar and Pitbladdo it is suggested that certain allocation mechanisms could adequately mimic the process of making the right decision. In the case with capacity constraints, the constraints reduce total output and act analogously to raising the variable cost of production. Once again, allocated costs could heuristically estimate the difference between prices and variable costs (inclusive of internal capacity prices).

**General Analysis**

Consider an industry where firms can make investments in productive capacity required for production of a certain group of products. Assume that capacity is available in fixed increments of size K, and that unit capacity costs are given by f, so that each capacity increment costs an amount fK to acquire. Direct costs, or operating costs, can vary from product to product, but unit operating costs are assumed to be unaffected by volume. Thus, there are no economies of scale or scope at the operating level. This assumption is made so that we may focus on use of common capacity as the factor linking products together. For the same reason, we shall assume that the demands for the products in the group we are considering are independent of one another and of the demands for other products.
The analysis will be of the single-period case, which admits two distinct interpretations. The first interpretation is that the analysis applies to a stationary environment where the fixed costs are incurred periodically. The second interpretation also involves a stationary environment, but with a one-time initial expenditure, with successive periods being capitalized into one. We are particularly concerned with how limited capacity is rationed among the various products. Firms must first decide how many capacity increments to acquire. Later, they must choose how much of each product to produce, with the knowledge of the industry-wide distribution of capacity. The price in each market is determined by the total quantity of that product produced, according to the (independent) demand curve for that product. That is, the price is what the market will bear at the aggregate production level. In game-theory parlance, there is an entry game where firms choose capacity, followed by Cournot competition (see Cournot, 1838; Kreps and Scheinkman, 1983).

In the latter stage, where operational decisions are made, firms observe who have entered and make assumptions about the quantity of each product the other firms will produce. Under these assumptions, but without hard evidence of what the others' production plans are, they make their own decisions of how much to produce. The price in each market is determined by the total quantity of that product produced, according to the (independent) demand curve for that product. That is, the price is what the market will bear at the aggregate production level. An equilibrium to this stage is characterized by equality between the production levels assumed by the firms and the production levels chosen under such assumptions. That is, equilibrium (which here is conditional on the distribution of acquired capacity) is a requirement for internal consistency. Any other prediction of the market outcome would contradict itself, if for some reason firms used such a prediction to form the basis of their assumptions regarding others' production levels.

The analysis of the former stage of capacity acquisition is sensitive to the order in which firms make such decisions and the information each firm has about the decisions made by other firms. For example, suppose a firm is able to make irreversible capacity decisions before the others, and the other firms will know what the first firm has committed to when they are formulating their own plans. Then the first firm will tend to exploit this by obtaining more capacity than it otherwise
would have, so that acquisition by competitors is more limited. In contrast to this scenario, we place firms on more of an equal footing. In particular, we assume that no firm can commit to a capacity strategy sooner than any other firm.

Although we allow an unlimited number of potential entrants, the cost increment $fK$ clearly limits the number of firms with positive capacity. The first question we ask is what the distribution of capacity will look like. Will every firm that installs some capacity install only one unit, or will there be a lesser number of firms, each installing multiple units? Due to the discreteness of the capacity increments, there may be multiple equilibria with different capacity distributions. But there is always an equilibrium in which each firm chooses only one increment or none at all. A verbal explanation of this is as follows:

As additional capacity is installed, industry-wide profit margins are reduced (or remain the same), since more goods are subsequently brought to market. In deciding to add another capacity increment, the additional revenue resulting from the production on the new unit must be balanced against not only the acquisition cost, but the decrease in revenue on existing units, due to the reduction in prices needed to accommodate the increased supply of products to the marketplace. It follows that the firms which benefit most from installing an additional unit of capacity are those with no existing volume to protect from decreases in profit margin. Since firms will add capacity until there is no net benefit from doing so, this argues in favor of active firms having only one increment of capacity (when capacity increments are very small relative to the size of the market, the above effect becomes small, and thus could be dominated by factors outside of our immediate consideration).

The complete characterization of the entire set of equilibria is well beyond the scope of this paper. Suffice it to say that the existence of equilibria for which some firms invest in more than one increment is likely to depend on parameters in an erratic manner that does not lend itself well to further analysis. Thus we shall assume firms invest in only one capacity increment, or stay out altogether. The questions we wish to address directly are how many firms should enter, and in what proportions should they allocate productive capacity to the various products? In addition, we
wish to relate these optimal decisions to systems such as internal pricing and overhead allocation to profit centers.

Let us first examine the question of allocating capacity. Recall that a firm must make assumptions about the production levels in the rest of the industry. Such assumptions can be linked with the demand curves to determine the market prices associated with each level of the firm's production. The 2-product case is represented in Figure 1. The "residual demand curves" denoted by $D_1$ and $D_2$, embody the price-quantity tradeoffs facing the firm, *given the assumed quantity of all the other firms*. Note that the first two vertical axes begin at the respective variable costs of the two products, denoted by $v$. If production levels are given by $AH$, then operating profits will be the areas $ACFH$. The change in revenue per unit increase in output is given by the marginal revenue curves $MR$. We shall assume without loss of generality that the product quantities are measured according to their use of capacity.

Figure 1: Residual demand, and marginal operating profit curves (marginal revenue less operating costs) for individual products (a, b) and in total (c).
Optimal allocation of capacity requires marginal revenue net of unit operating cost (marginal operating profits) to be equalized over all products with positive production. This happens to be true of the quantities A1H1 and A2H2, where marginal revenue net of unit operating cost is A1B1 = A2B2. The other requirement for an optimal production plan is that all available capacity is fully utilized unless marginal revenue equals unit operating cost for each product. Figure 1c depicts the horizontal sum of the marginal revenue curves in 1a and 1b. For example, B1G1 + B2G2 = NR. Thus if total capacity were NR = LS, then the production plan (A1H1,A2H2) would be optimal. If capacity were LT or greater, then the optimal production plan would be (A1J1,A2J2), the common marginal operating profits being zero. Thus figure 1c identifies the marginal operating profit for both products as a function of capacity K. This value is also the shadow price for an infinitesimal increase in capacity (a fact that is not particularly useful in our context, since capacity increments are so large relative to firm output).

Now suppose the firm were organized by product-specific profit centers, with the capacitated common resource being a service center. Production quantities and prices for each product are determined by that product's manager. An internal price for use of the service center by each of the products is set by central administration, with each product manager having the incentive to maximize profits net of internal payments. If central administration has access to the information embodied in figure 1c, it can set the internal price accordingly. However, to literally construct figure 1c, one needs to know the marginal operating profit functions of each of the profit centers. But with such information in the hands of central administration, optimal production quantities can be easily computed centrally, so it is not all that clear why product managers should control such decisions in the first place.

The answer to this question thus hinges on whether there are shortcuts to estimating the appropriate shadow price that do not require such detailed knowledge. For example, one method used in practice is to select an internal price of unit capacity costs f. This is equivalent to full cost allocation of the capacity costs when all of capacity is utilized. Service center revenues would just cover its fixed costs.
Returning to Figure 1, suppose capacity is given by LS. Then operating profits from each product are given by areas ACFH and center revenue is given by area LNRS, which is also the sum of areas BCFG. If the shadow price happens to equal unit capacity costs, NL = f, then operating profits exceed fixed cost by an amount equal to the sum of the areas BCFG, the sum of profit center net profits. Using Figure 1 alone as a guide, the occurrence f=NL would appear to be mere happenstance, as capacity costs do not directly enter the picture. However, the location of the residual demand curves D facing the firm depend critically on how many firms have entered, which in turn depends on capacity costs.

Indeed, with unrestricted entry of identical firms, equilibrium would require erasure of net profits if but one more firm were to enter. In order for net profit margins to be significantly positive, therefore, the addition of an additional increment of capacity must have a significant repercussion in the marketplace.

In order to examine intermediate cases from a graphical perspective, we shall consider the single product case. While this case is not particularly well motivated from a practical perspective, the resulting insights do carry over to multiproduct situations. Firms producing at capacity will make positive profits only when the market price exceeds total unit costs, i.e. \( p > v + f \). Figure 2 enables us to determine the number of firms consistent with an equilibrium where each firm is producing at capacity \( K \). The demand curve \( D \) is the overall market demand curve, rather than a single firm's residual demand curve, the horizontal scale being the total quantity supplied to the market by all firms.
We see that if there are four firms or less, market price will exceed v+f, but not if there are five firms or more. There would thus have to be four firms investing in capacity K in order for there to be an equilibrium (without excess capacity).

Figure 3 is magnification of the upper right corner of figure 2, representing a translation of the residual demand curve facing each firm (Note that marginal revenue crosses the 4K vertical line above the horizontal axis, so that the firm's marginal revenue exceeds unit operating costs at full capacity. This ratifies the assumption that each firm chooses to operate at full capacity in equilibrium.). The optimal internal price λ (the shadow price of capacity) is the height of the intersection between MR and the 4K line. A product manager maximizing revenue net of direct variable cost and internal charges qλ, but ignorant of the capacity limitation, will nonetheless choose to exactly exhaust capacity. In this sense, the scarcity of the capacitated resource is decentralized via the internal price λ in a way to balance demand for the resource with its supply.
With linear demand, the marginal revenue curve is also linear and its slope is twice as steep. As a result, MR crosses the 4K line exactly at the same level where the demand curve crosses the 5K line, whereby the shadow price is equal to the operating profit margin \( p-v \) that would prevail if one too many firms entered. We know that the latter quantity is less than unit capacity costs \( f \). Thus with linear demand we know that the optimal internal price never exceeds \( f \) (this will hold in general except for some cases where demand bows outward to the upper right).

![Figure 3: Marginal revenue and residual demand curves for an individual firm.]

Note that if an internal price of \( f \) were employed, the product manager would choose to produce at less than full capacity. Thus charging unit capacity costs \( f \) is unambiguously inconsistent with recovery of the capacity cost through internal revenues. As a consequence of the entry process, full cost allocation via the internal price would necessarily result in excess capacity, regardless of the magnitude of the capacity constraint. In order to fully allocate fixed costs, the internal price would have to increase further to make up for the lost revenue due to capacity underutilization (causing further underutilization). This would work only if there was sufficient
room for profits in the first place, and if the marginal revenue curve is sufficiently steep so that the increases in the internal price can overcome declining utilization (A minimal requirement is that $bK > 2f$, where $b$ is the slope of the demand curve).

Figures 2 and 3 represent somewhat of an intermediate case between the extremes where there is almost room for one more firm and where net margins are near zero for existing firms. Figure 4 represents these extreme cases, respectively. Starting with four firms in figure 4a, we see that an additional firm would bring profits just to zero, so there is nearly an incentive for another firm to enter. Note that the optimal internal price actually equals unit capacity costs. In figure 4b, on the other hand, four firms make just enough operating profits to cover their fixed cost. The intercept of the residual demand curve is $f+bK$, where $b$ is the slope of the demand curve. Since the slope of MR is $2b$, MR crosses the 4K line at $(f+bK) - 2bK = f-bK$. 
Figure 4:  (a) A case where an additional firm could enter and just cover costs;
(b) the four incumbents just cover their costs.

Thus for the linear approximation of demand, shadow price is between $f$ and $f-bK$,
approaching each extreme in the cases where there is almost room for another entrant and where
there is just enough room for the final entrant, respectively. This supports our intuitive analysis
that $f$ is a good approximation for the internal price when the capacity increment $K$ is small. As $K$
increases with $f$ remaining fixed, the fluctuations between the two bounds become larger and larger until full capacity utilization is not sustained in equilibrium. Figure 5 displays the exact form of the relation between shadow price and capacity increment, for a fixed value of $f$, under a special assumption. Each line segment corresponds to a given number of firms, with the jumps occurring at the borderline between entry and exit. Figure 6 presents the general (complementary) case. Detailed deviations are given in Appendix A.

These results, while interesting, are perhaps somewhat disappointing from the viewpoint of practical application. Except for the endpoints where $K$ is small and large, there is no simple rule enabling central administration to set an internal price as a routine function of internal accounting data. Instead, more detailed direct knowledge of the competitive environment is called for. The unequivocal qualitative insight, however, is that under these circumstances full cost allocation typically overcharges for utilization of a scarce resource. This is true particularly when large efficient scale results in firms' operating decisions having a nontrivial impact on the marketplace.

We have seen that a firm which tries full cost allocation while others are profit maximizing faces a difficult if not impossible task. Conversely, if full cost allocation based on service utilization is an industry standard, there may be significant profit opportunities for firms who are willing to break from such a practice. Suppose active firms price internally to cover fixed costs, and other potential entrants will stay out if it is not possible for internal revenue to cover fixed costs. Then total firm profits exceed fixed costs at least by the net revenue accruing to the profit centers. In addition, active firms' production levels will be less than optimal unless they are at capacity. These two factors may leave a window of opportunity for a profit maximizing entrant, even when the door is shut to firms who insist on covering fixed costs with internal revenue.

If a firm wishes to maximize profits, but also wants to remain within the framework of full cost allocation, it is clear that the fixed capacity costs should be allocated to products on a basis different than utilization of the common capacity. Extending Karmarkar and Pitbladdo (1990), the appropriate thing to do would be to allocate the remainder of the fixed costs (after internal payments are collected) on the basis of profit center net revenues (gross margin). This would have
the additional advantage that no profit center will show a loss unless it was a mistake to acquire the capacity.

Figure 5: The shadow price (λ) as a function of this size of the capacity increment K for the case \( S = \alpha^2/\beta \).

Figure 6: The shadow price (λ) as a function of this size of the capacity increment K for the case \( S^2 > \alpha^2/\beta \).

19
Discussion and Extensions

Even though the cost of capacity, at the time that firms make production decisions, is a sunk cost, it is nonetheless a major determinant of the internal price that would optimally allocate capacity. The mechanism by which the cost of capacity affects the optimal internal price is in the way it determines the capacity decisions of all the firms in the market, which in turn affects the extent of competition a firm faces at the time production decisions are made. In the context of homogeneous firms operating in a stable, deterministic environment, we have seen that the optimal internal price is unit fixed cost when capacity increments are small, zero when capacity increments are large, and somewhere in between but not easily captured by accounting data for intermediate cases.

We have assumed that capacity is obtained in fixed increments, and that there is a definite capacity limit. In many situations, capacity limitations do not act as a hard constraint, but instead cause variable costs to increase, so that complete utilization of capacity is discouraged. These costs may have diverse underlying causes, including increasing inefficiency in production, the need to use more costly capacity alternatives including overtime, queueing delays and increased lead times, and externalities such as quality problems and increased downstream safety stock requirements. For example, Karmarkar (1987) points out that work-in-process and lead time costs effectively substitute for capacity constraints when queueing delays are present; Banker, Datar and Kekre (1988) elaborate on the costs of capacity in this situation; Magee (1956), and Karmarkar and Lele (1989) model the safety stock consequences of the increased lead times created by high capacity loading. There do not appear to be many studies investigating the allocation of costs in such cases. Zimmerman (1979) has suggested that cost allocations might be a way to convey the externalities due to congestion and delays. Miller and Buckman (1987) have used a multi-server queueing model to examine the allocation of fixed costs in the presence of congestion and increasing marginal costs of capacity utilization. They conclude that when capacity costs are linear, the marginal variable cost is a good approximation, but that this is not the case when there are
significant economies of scale. Questions of entry and competition are not addressed in these papers.

When such increasing marginal costs become the limiting factor on production quantity and the nominal capacity constraint is never binding, the question arises as to whether the results of the preceding analysis hold at a qualitative level. For example, is the "right" internal price of capacity now determined by endogenous data? Is this price related to the cost of purchasing capacity increments? The answer to the first question is a qualified yes. While the internal price is related to the marginal variable cost (which is presumed to be internally known), it also depends as before on the number of firms in the market. Furthermore, as before, this price will not generally cover the fixed costs of capacity acquisition. Nor is there support for Zimmerman's conjecture regarding congestion externalities. The marginal variable costs of congestion are bounded by the fixed costs of capacity acquisition, but are not otherwise determined by them in a simple way. The detailed analysis is very similar to the uncapacitated case of Karmarkar and Pitbladdo (1990), with the constant variable cost replaced by the marginal variable cost rate; it is described in Appendix B.

The assumption of homogeneous firms played an extensive role in the analysis. Suppose instead that firms differ in their capacity cost. If firms were restricted in some way to a single capacity increment, then the analysis goes through with little change, except that each active firm's capacity cost would be replaced by the smallest capacity cost of active firms, since it is the latter that determines the entry threshold. If there were no such restrictions, then firms with lower costs would invest in more capacity increments, until the point where prospective marginal revenue falls to the vicinity of unit costs, accounting for integer effects. In this case, it is the firm's own cost that is germane to the optimal internal price. Furthermore, the qualitative results still hold: the optimal internal price is unit fixed cost when capacity increments are small, zero when capacity increments are large, and somewhere in between but not easily captured by accounting data for intermediate cases.

If firms make capacity decisions in the face of uncertainty over future economic conditions, the analysis is fundamentally altered. Production decisions can be based on more current information.
regarding economic conditions, and the internal price must be adjusted accordingly. For example, if demand turns out to be lower than expected for the majority of the products, there will be lower demand for capacity at any given internal price. This if the internal price is based on conditions that had been expected to prevail, rather than those that actually prevail, there may be inefficient excess capacity, bearing in mind that the capacity costs are sunk.

Our model involves a single period for capacity acquisition and a single period for production. It is trivially extended to multiple periods for production, provided there is no uncertainty, seasonality, or other source of intertemporal variability. The single period of capacity acquisition glosses over the issue of whether capacity costs should be historical, current, or "opportunity costs". Opportunity costs based on, for example, liquidation value ignore the legacy of the capacity costs in determining the number of firms in the marketplace. If current capacity costs have risen to the extent that further entry has been eliminated, it is the historical costs that have determined the number of firms, and hence have the most relevance to the optimal internal price. On the other hand, if there is continued entry at the current prices, it is the current prices that bear on the number of firms, and hence the optimal internal price. Whereas a precise statement in this regard can be made only with an explicit multiperiod analysis, the point is that there is no reason to automatically discard historical costs as being inappropriate to current conditions.

Gilbert and Harris (1984) study the dynamic problem of lumpy investment in a growing market for the single product case. They consider closed loop (perfect) equilibria to a multifirm game of capacity expansion. Equilibria consist of an initial stage that is has similar qualities to our analysis, but as time progresses and firms compete over the timing of capacity expansion, their strategies are more aggressive. Firms expand before the increment is profitable in order to prevent others from preempting them. The investment will turn profits before another expansion occurs, but the net present value over the cycle is competed away to zero. The implication for our analysis is that in such a game, the internal price is expected to be even lower, with our lower bound becoming the norm rather than the extreme, underscoring the observation that unit fixed costs overestimate the internal price.
References


Jaikumar, R. - to be added


Appendix A: Analytical Derivations

For simplicity, assume that in the relevant range the market price $p_i$ for each good depends linearly on $Q_i$, the aggregate quantity supplied:

$$\begin{align*}
    p_i &= a_i - b_i Q_i \\
\end{align*} \quad (1A)$$

This should be viewed as a linear approximation rather than a restriction. Each firm that has invested $fK$ in production capacity must choose a quantity $q_i$ to supply each market $i$, taking the production plan of others as given. The firm's operating profit in market $i$ is:

$$\begin{align*}
    (p_i - v_i) q_i &= (a_i - b_i Q_i - v_i) q_i \\
    &= [a_i - b_i (Q_i - q_i) - b_i q_i - v_i] q_i \\
\end{align*} \quad (2A)$$

where $v_i$ represents the firm's variable cost per unit of product $i$. $Q_i - q_i$ is the total production of product $i$ by others, which the firm takes as given.

Thus the firm seeks to maximize

$$\sum_{i} [a_i - b_i (Q_i - q_i) - b_i q_i - v_i] q_i \quad (3A)$$

subject to the constraint that

$$\sum_{i} t_i q_i \leq K \quad (4A)$$

where $t_i$ represents the amount of capacity required to produced one unit of product $i$.

Thus for some $\lambda \geq 0$, the solution to the above problem is
or zero, whichever is greater. If \( n \) identical firms have acquired the capacity \( K \), then \( Q_i = nq_i \), and (5A) becomes

\[
q_i = \frac{a_i - b_i (Q_i - q_i) - v_i - \lambda t_i}{2 b_i}
\]

For simplicity, we shall hereafter presuppose the case where all products are produced in positive quantities.

With excess capacity, \( \lambda = 0 \) and

\[
q_i = \frac{a_i - v_i}{(n + 1) b_i}
\]

Thus the condition for excess capacity is:

\[
\sum_i t_i \left( \frac{a_i - v_i}{(n + 1) b_i} \right) \leq K \quad (8A)
\]

If (8A) is not satisfied, \( \lambda \) is found by equating

\[
\sum_i t_i \left( \frac{a_i - v_i - t_i \lambda}{(n + 1) b_i} \right) = K \quad (9A)
\]

whereby
\[ \lambda = \alpha - \beta (n + 1) K \]

where \( \beta = \left( \sum_i t_i^2/b_i \right)^{-1} \)  \hspace{1cm} (10A)

\[ \alpha = \beta \sum_i \frac{t_i}{b_i} (a_i - v_i) \]

Note that \( \alpha \) and \( \beta \) are independent of \( n \) and \( K \). We see that \( \lambda \), representing the shadow price of capacity, decreases as expected with the number of firms and with the size of the capacity increment, until the point where excess capacity prevails. Substituting (10A) into (6A) we have:

\[ q_i = \frac{a_i - v_i}{(n + 1) b_i} - \frac{t_i \alpha}{(n + 1) b_i} + \frac{t_i}{b_i} \beta K \]  \hspace{1cm} (11A)

We see that, as the capacity increment increases, while holding the number of firms constant, production increases across the board, but particularly for those products with high capacity requirement \( t_i \) relative to price sensitivity \( b_i \). This trend continues until the third term catches up with the second term, whereupon excess capacity prevails.

As the number of firms increases, production is shifted away from those products whose capacity requirement \( t_i \) is small compared to \( (a_i - v_i) \), the excess of the demand intercept over variable cost, toward those products whose capacity requirement is high relative to potential profit margin. Again, this trend continues until excess capacity prevails.

Conditional on the number of firms \( n \) acquiring the capacity \( K \), the market prices are given as:

\[ p_i = a_i - b_i n q_i \]
Operating profits in market i for each firm are given as:

\[
(p_i - v_i) q_i = \left( \frac{a_i - v_i + n t_i \lambda}{n + 1} \right) \left( \frac{a_i - v_i - t_i \lambda}{b_i (n + 1)} \right)
\]  

(13A)

Summing to get total profits:

\[
\sum_i (p_i - v_i) q_i = \sum_i \left( \frac{(a_i - v_i)^2 + (n - 1) t_i \lambda (q_i - v_i) - n t_i^2 \lambda^2}{b_i (n + 1)^2} \right)
\]

\[
= (n + 1)^{-2} (S + (n - 1) \lambda \alpha / \beta - n \lambda^2 / \beta)
\]

(14A)

where \( S = \sum_i \frac{(a_i - v_i)^2}{b_i} \)

Substituting \( \lambda = \alpha - \beta (n + 1) K \), and regrouping terms, we get an expression for profit as a function of the number of firms \( n \) and the capacity increment \( K \):

\[
\Pi(n, K) = \frac{S - \alpha^2 / \beta}{(n + 1)^2} + (\alpha - n \beta K) K
\]

(15A)

In order to demonstrate that operating profit \( \Pi(n, K) \) decreases as the number of firms increase, note that by the Cauchy-Schwartz inequality:
By squaring each side we see that $S \geq \alpha^2/\beta$. Note that equality is attained if and only if the ratios $(a_i - v_i)/t_i$ are independent of $i$ (whereupon the common ratios are equal to $\alpha$).

The number of firms acquiring capacity $K$ will be determined by comparing operating profits $\Pi(n, K)$ with acquisition cost $fK$. For firms to acquire $K$, it must be that the operating profits cover the cost of acquisition, $\Pi(n, K) \geq fK$. On the other hand, in order for there to be no room for further entry, it must be the case that the addition of one more firm will cause operating profits to fall to the extent that they are no larger than acquisition costs: $\Pi(n + 1, K) \leq fK$. Thus, if we let $x^*$ be the real number which solves

$$fK = \Pi(x, K) \quad (17A)$$

then the equilibrium number of firms, $n^*$, is equal to $[x^*]$, the greatest integer in $x^*$ (when $x^*$ is an integer, then $n^*$ can take on two values, $x^*$ and $x^* - 1$).

From (14A) and (15A), we see that $x^*$ solves
\[
f_{K} = \frac{S - \alpha^2/\beta}{(x + 1)^2} + (\alpha - x\beta)K \tag{18A}
\]

Unless \( S = \alpha^2/\beta \), equation (18) is cubic in \( x \). As such, the closed-form expression for \( x^* \) is quite complex. Thus we shall examine the case \( S = \alpha^2/\beta \) and then make qualitative adjustments for the general case.

Recall that \( S = \alpha^2/\beta \) when \( a_i - v_i \) are uniformly proportional to \( t_i \), whereupon \( a_i - v_i = \alpha t_i \). In this case, optimal production quantities in capacity constrained cases are uniformly proportional to those in the unconstrained case; they are simply scaled back (see equation (6)). Otherwise, when \( S > \alpha^2/\beta \), firms increase the proportion of production devoted to products where \( t_i \) is small compared to \( a_i - v_i \). In fact, the second term of the RHS of (15A) represents the profits if firms simply scale back production from the optimal unconstrained levels. The first term represents the extra profit attained because firms change their production proportions.

When \( S = \alpha^2/\beta \), (18A) is readily solved:

\[
x^* = \frac{\alpha - f}{\beta K} \tag{19A}
\]

(19A) also represents a lower bound for \( x^* \) in the cases \( S > \alpha^2/\beta \).

We can now derive expressions for the Lagrange multiplier, taking into account the endogeneity of the number of firms.

\[
\lambda = \alpha - \beta(n^* + 1)K = \alpha - \beta \left[ \frac{\alpha - f + \beta K}{\beta K} \right] K \tag{20A}
\]

Since \( x^* \geq n^* \), a lower bound for \( \lambda \) is
\[ \lambda \geq \alpha - \beta(x^* + 1)K = f - \beta K \]  \hspace{1cm} (21A)

and this lower bound is approached as \( x^* \) is slightly larger than integral, representing the case where there was just enough room for the last firm to enter, and operating profits barely cover fixed cost. Since \( x^* - 1 \leq n^* \) an upper bound for \( \lambda \) is:

\[ \lambda \leq \alpha - \beta x^* K = f \]  \hspace{1cm} (22A)

and this lower bound is approached as \( x^* \) is slightly smaller than integral, representing the case where there is almost room for another firm to enter.

Thus the optimal internal price for operating decisions can fluctuate by the amount \( \beta K \) on either side of the marginal cases, when firms are capacity constrained. With these limits in mind, figure 5 represents equation (20A). The number of firms decreases in steps as \( K \) increases. When \( K \) is small, there are many firms and the segments have very steep slope but have little range, always being close to \( f \). As \( K \) increases, the segments are more shallow but have a larger range.

Returning to the general case where \( S > \alpha^2 / \beta \), the linear segments remain the same because it is still the case that \( \lambda = \alpha - \beta(n + 1)K \). However, the jumps will be displaced to the right, as in figure 6. This is because (19A) underestimates \( x^* \). It turns out that the jumps are still of size \( bK \), as is easily verified.
Appendix B: The Case of Increasing Variable Costs

Suppose that the variable costs of production, given production quantities \( \{q_i\} \), are given by a function \( V(q) \) that may not be separable by item. We also assume that the variable cost function is such that the capacity limit is never reached; i.e. the cost function rises sufficiently quickly as production quantities approach the capacity limit. The analysis proceeds similarly to the capacitated case, except that there is no Lagrange multiplier due to the constraint, and the variable cost is replaced by the rate of change of the variable cost function. Thus the firm's total operating profit, is given by an equation analogous to (2)

\[
\sum_i p_i q_i - V(q) = \sum_i [a_i - b_i (Q_i - q_i) - b_i q_i] q_i - V(q)
\]

(1B)

The firm seeks to maximize the sum of the operating profits in all markets. Given \( n \) identical firms in the market, \( Q_i = nq_i \), and the condition for maximization can be written as

\[
a_i - (n+1)b_i q_i = \nabla_i V(q), \quad \forall i
\]

(2B)

where \( \nabla_i V(q) \) is the first derivative of the variable cost function with respect to \( q_i \).

Equivalently, the production quantity for item \( i \) solves

\[
q_i = \frac{q_i - \nabla_i V(q)}{(n+1)b_i}
\]

(3B)

Note that if \( V \) is a convex function, the conditions for profit maximization can be interpreted as the stationarity conditions for a convex minimization problem, with the objective function bounded below. Essentially, this shows that a solution exists, and under mild conditions is unique.

The price in market \( i \) conditional on there being \( n \) firms in the market is
The total profits for each firm contingent on the number of firms in the market are given by

\[ R(n) = \sum_i p_i q_i - V(q) \]  

(5B)

At the entry decision, the number of firms entering will be \( n \), such that

\[ R(n) \geq F \geq R(n+1) \]  

(6B)

For specific cases of the function \( V(q) \), the optimal production quantities can be determined given \( n \), and the function \( R(n) \) calculated for values of \( n \). The last condition then gives the number of firms entering. Unfortunately, it appears to be difficult to obtain these results in closed form.

We consider two examples of choices for \( V \):

**Example 1**: Suppose that

\[ V(q) = \sum_i (u_i q_i + (w_i/2K)q_i^2) \]

where \( K \) represents the capacity increment obtained at a fixed cost \( F \). It is assumed that the parameters are such that the capacity limit is not exceeded; i.e. the marginal costs rise sufficiently quickly. This separable form does not capture product interactions due to contention for capacity, but does have the property that variable costs increase with production levels. In this case,

\[ \nabla_i V(q) = u_i + (w_i/K)q_i \]

Using this in the condition (2B) gives the optimal production quantity

\[ q_i^* = (a_i - u_i)/[(n+1)b_i + (w_i/K)] \]

This expression is similar to the uncapacitated case in Karmarkar and Pitbladdo (1990), except for the second term in the denominator. Note that this term acts as though increasing the elasticity of
the demand function, and leads to a smaller production quantity. For this case, the expression for total profit can be written in closed form as

\[ R(n) = \sum_i (a_i - u_i)^2 (2 + (w_i/b_iK))/2b_i[n+1 + (w_i/b_iK)] \]

As \( K \) goes to \( \infty \) or as the \( w_i \) go to 0, this expression simplifies to the uncapacitated case. The number of firms in the market can be found numerically by locating \( n^* \) such that

\[ R(n^*) \geq F > R(n^*+1) \]

**Example 2:** Consider the case

\[ V(q) = \sum_i u_i q_i + \sum_i w_i[D/(K-D)] \]

where \( D = \sum_i t_i q_i \), the total usage of capacity, and \( w_i \) is a weighting factor for the \( i \)'th item. In this case, the marginal cost rate for item \( i \) can be written as

\[ V_i V = u_i + (t_i/D)V(1 + V) \]

No further simplification seems to be forthcoming. The production quantities can be computed numerically for a given \( n \), and then the condition for entry tested for increasing \( n \) till the number of firms entering is determined.

In the case of increasing variable costs, the equivalent of the Lagrange multiplier term \( \lambda t_i \) of the earlier analysis, is the marginal cost rate \( V_i V(q) \). This cost rate can indeed be determined from endogenous data about the characteristics of the manufacturing system. However, to determine the correct production quantities, it is necessary to have information regarding the number of firms in the market place, since these quantities must satisfy equation (2B). From another viewpoint the conditions in (2B) suggest how the process of determining production quantities can be
decentralized. Suppose that each product \( i \) is managed by a product manager. Note that the left side of each condition is \((p_i - bjq_i)\), which is the marginal revenue from an additional unit of production of item \( i \). Assuming that this information is known to product managers, the centralized information required is just the marginal cost rate. In fact in the separable quadratic case, this can also be decomposed by item. In the general case, decentralization could be achieved by iterating between trial production quantities declared by product managers, and the marginal cost rates due to resource usage, calculated centrally from these quantities. Furthermore, the marginal cost rate does not "cover fixed costs". As in the earlier analysis of this paper, the fixed costs are relevant primarily at the entry decision, and must be recovered from the excess contribution margin by which revenue exceeds variable costs.