

NBER WORKING PAPERS SERIES

THE THEORY OF ALLOCATION AND ITS IMPLICATIONS
FOR MARKETING AND INDUSTRIAL STRUCTURE

Dennis W. Carlton

Working Paper No. 3786

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 1991

I thank NSF and the program in Law and Economics at the University of Chicago for financial assistance. I also thank K. Crocker, E. Lazear, S. Peltzman, A. Shleifer, G. Stigler, and participants at seminars at the Department of Justice, MIT, NBER, Northwestern, Pennsylvania State, Princeton, University of Chicago and University of Florida. This paper is part of NBER's research programs in Economic Fluctuations and Industrial Organization. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Working Paper #3786
July 1991

THE THEORY OF ALLOCATION AND ITS IMPLICATIONS
FOR MARKETING AND INDUSTRIAL STRUCTURE

ABSTRACT

This paper identifies a cost of using the price system and from that develops a general theory of allocation. The theory explains why a buyer's stochastic purchasing behavior matters to a seller. This leads to a theory of optimal customer mix much akin to the theory of optimal portfolio composition. It is the job of a firm's marketing department to put together this optimal customer mix.

A dynamic pattern of pricing related to Ramsey pricing emerges as the efficient pricing structure. Price no longer equals marginal cost and is no longer the sole mechanism used to allocate goods. It is optimal for long term relationships to emerge between buyers and sellers and for sellers to use their knowledge about buyers to ration goods during periods when demand is high. This rationing can take the form of refusing to sell to new customers and putting established customers on quotas. The evidence shows that this form of rationing, though foreign to the thinking of most economists, characterizes several industries.

The theory provides an important incentive for a firm to exist, namely to facilitate trade amongst its customers. The theory also provides a convincing explanation for the hostility that new futures markets face from established firms in the industry and shows that several practices, like price differences amongst consumers and swapping product with rivals, can be the result of competition and not market power.

Dennis W. Carlton
Graduate School
of Business
University of Chicago
1101 East 58th Street
Chicago, IL 60637
and NBER

I. Introduction

Microeconomics is concerned with the efficient allocation of goods. It is typical for economists to focus on the price system as the mechanism used to achieve this efficient allocation. An impersonal auction system is often in the back of most economists' minds when they think of efficient resource allocation. Other mechanisms to allocate goods such as rationing are usually considered inferior to a price system and are not given much attention.¹ The economist's exclusive focus on price comes from an implicit assumption that use of the price system is costless. After all, if it is costless to use the price system to achieve an efficient allocation, why not use it?

This paper explains that economists' faith in the price system is misplaced from the viewpoint of both theory and evidence. Prices alone should not be viewed as the most efficient way of allocating goods in many industries. The paper identifies a cost of using the price system and from that develops a general theory of allocation that has significant implications for optimal allocation and industrial organization. The theory explains the role of market making and marketing -- and thereby provides an explanation for important features of the modern corporation, such as

1. Coase (1937) and Weitzman (1974) are notable exceptions.

marketing and planning departments, that have received relatively little attention from economists.

The theory explains how even the most physically homogeneous good can be quite heterogeneous from the seller's vantage point depending on the buyer's stochastic purchasing characteristics. This leads to a theory of optimal customer mix much akin to the theory of optimal portfolio composition. It is the job of a firm's marketing department to put together this optimal customer mix. A dynamic pattern of pricing related to optimal Ramsey pricing emerges as the efficient pricing structure. Price no longer equals marginal cost and is no longer the sole mechanism used to allocate goods. It is natural and optimal for long term relationships to emerge between buyers and sellers and for sellers to use their knowledge about buyers to ration goods during periods when demand is high. This rationing can take the form of refusing to sell to new customers and putting established customers on quotas. This form of rationing, though foreign to the thinking of most economists, seems to characterize several industries, as I report later on. The theory provides an important incentive for a firm to exist, namely to facilitate trade amongst its customers. The theory identifies the incentives that firms have to invest in marketing and planning. Finally, the theory provides a convincing explanation for the hostility that new futures markets face from established firms in the industry and shows that

several practices, like price differences to consumers and swapping product with rivals, that are often associated with market power can instead be the result of competitive market forces.

This paper is organized as follows: Section II explains why it is wrong to regard the use of the price system as free and presents some puzzling facts that do not accord with the economic intuition from price clearing theories of market behavior. Section III develops the theory of optimal allocation in the presence of a cost of using the price system under some simple assumptions. Sections IV, V and VI examine equilibrium and show how to modify the theory so as to relax several of its simple assumptions. Section VI explains how a firm can facilitate trade amongst its customers. Section VII reports on empirical evidence that supports the theory. Section VIII discusses the implications of the theory for the role of marketing, choice of customer mix, capital investment, attitudes toward the development of futures markets and swapping with rivals. Section IX presents conclusions.

II. Organizing A Market That Clears By Price is Costly, and For That Reason Many Markets Do Not Clear By Price

Every economics student learns that price adjusts so as to equate supply and demand. A (fictional) auctioneer often is used to explain how price is set. This auctioneer doesn't even earn the minimum wage. All firms are assumed to have access to the price mechanism that allows them to costlessly market their goods. There is no need for long term relationships between buyers and sellers and non-price rationing never occurs.

Most economists recognize that there are relatively few markets like the one I just described, yet this description is the model that guides most economists' intuition. I will soon describe why that intuition can sometimes be misleading. Before I do that, it is worth emphasizing the costs of setting up an auction market based on the few actual examples we have of such markets. Once the costs of setting up an auction market are recognized, it follows immediately that other allocation mechanisms may be superior.

One can learn a great deal about the cost of a price system by studying institutions whose business is to create such markets (see Carlton (1984))². For example, if one studies the Chicago Board of Trade one observes the following:

1. a large building in an expensive part of town,

2. Carlton, D.W., "Futures Markets: Their Purpose, Their History, Their Growth, Their Successes & Failures," Journal of Futures Markets, (Fall 1984) 4(3):237-271.

2. many people involved in each transaction (e.g. brokers, floor traders, employees of a clearing-house, consultants),
3. office buildings housing the people involved in each transaction, and
4. elaborate record keeping.

Undoubtedly the greatest cost is the time cost of all the people involved. A significant fraction of the economy of the City of Chicago is devoted to the making of markets. If a magic spell could be cast to make transactions costless, the Chicago economy would be devastated, at least in the short run. This emphasizes how far from costless the making of markets really is.

Aside from the fact that there are costs to setting up such markets, it is also true that many markets once created fail. Roughly one half of all successful futures markets fail within 10 years of their introduction. (See Carlton (1984))³. Since markets clearly produce at least some benefits to its creators, the high failure rate must indicate the presence of significant costs of operation which outweigh these benefits. Furthermore, only a small fraction of products have ever had futures markets. Again, this emphasizes that there must be significant costs to the creation of such markets, otherwise the benefits flowing from such markets would suggest a proliferation of such markets.

3. See note 2, supra at 5.

An auction market generates positive externalities. It provides valuable information to non-participants without charge. But once it is recognized that it takes real resources to create a market, the existence of the positive externality guarantees that there is no reason to expect the efficient number of markets to be created.⁴ Those who actually transact in such a market may be charged a brokerage fee, or be faced with a bid-ask spread, or with asymmetric information that allows brokers to trade profitably. But the point remains that the marginal charge for participation may not reflect social benefits and therefore may fail to create incentives for the broker or market maker to continue the market even when the social benefits are great. Once it is recognized that "a market" produces a product—namely the allocation of the good—the conclusion that there is not an efficient incentive to create markets follows by analogy to the work on optimal product variety. See Spence (1976)⁵ and Dixit and Stiglitz (1977)⁶. If markets do not get created, alternative mechanisms to allocate goods must be used.

4. The existence of one market can affect the benefits of another. In such a setting, it is conceivable that either too many or too few markets get created.

5. Spence, M., "Product Differentiation & Welfare," American Economic Review, (May 1976) 407-414.

6. Dixit, A., & Stiglitz, J., "Monopolistic Competition and Optimum Product Diversity," Bell Journal, (1977), 297-308.

Aside from the relative paucity of auction type markets, there is considerable additional evidence that does not square with the predictions that emerge from the usual market clearing model. I have elaborated on some of this evidence at length elsewhere (Carlton (1986, 1989))⁷, and so only summarize the evidence here.

For what appear to be homogeneous goods,⁸ there is often a very low correlation in price movements across buyers. Moreover, buyers and sellers are often "married" to each other for long periods of time (e.g. several years). It is frequently true in some industries that price, once set to a buyer, does not change for periods in excess of one or two years or more. The degree of market concentration is related to the rigidity of price. There are frequent claims in some industries that non-price rationing is occurring when demand is high with regular customers receiving preferred treatment.⁹ Finally, the rise of the modern corporation in the late 1880's resulted in the need to have salesman and marketing departments whose responsibility was to

-
7. Carlton, D.W., "The Rigidity of Price," American Economic Review, (September 1986) and "The Theory and Facts of How Markets Clear: Is Industrial Organization Useful for Understanding Macroeconomics," Chapter 15 in R. Schmalensee and R. Willig ed., Handbook of Industrial Organization, (1989). A revised version appears in Chapter 21 of D. Carlton and J. Perloff, Modern Industrial Organization, Scott-Foresmen (1990).
 8. The goods are typically intermediate manufactured goods (e.g., steel, copper, paper and chemicals).
 9. Section V presents some of this evidence in detail.

learn about their customers and sell goods. (Chandler (1977))¹⁰. This central feature of the modern corporation is not easily explicable in a world in which goods can be costlessly sold on an auction market.

III. The Cost of Using the Price System

In the absence of an auction market, a firm that sets a price runs the risk that it has set the wrong price. If the price is too low, too much is demanded and the firm has to ration customers. If the price is set too high, then too few customers purchase and the firm is stuck with excess capacity. By the time the firm figures out whether it has set the wrong price, it is too late -- the customers have already gone (if the price is too high) or have placed their order at the low price (if the price is too low). This then creates a real cost of using the price system to allocate goods. More precisely, because of uncertainty as to the level of demand, the firm is unable to set the market clearing price. The information set over which the firm can set price ex ante is "coarser" than the ex post realizations of demand. (Later on, we see how this information set is endogenously determined).

10. Chandler, A., The Visible Hand, Harvard University Press, Cambridge, Mass (1977).

Let us illustrate this cost of using the price system in the context of a simple model. See Figure 1. Suppose that there is a firm whose production technology involves a fixed cost F and a constant marginal cost up to a fixed capacity K . Suppose that the customers of the firm in aggregate have a demand curve of either D_L (the low state of demand) or D_H (the high state of demand) but the firm does not know which of the two demand curves is applicable. If the firm knew that D_H prevailed, it would charge a market clearing price of λ_H , while if it knew that D_L prevailed, it would charge a market clearing price of λ_L .¹¹ The firm can, by assumption, choose only one price. No matter which price the firm chooses, there will be an allocative inefficiency. If it charges a price above λ_L , it creates a deadweight loss when D_L prevails. If it charges a price below λ_H , it is likely to create a deadweight loss when D_H prevails since it faces excess demand. For example, as Figure 1 illustrates, if the firm charges λ_H , the firm would create a zero deadweight loss when D_H prevailed, but would create a deadweight loss of ABCD when D_L prevailed.

11. I will soon focus only on firms that compete with each other to attract customers. The statement in the text also holds true for a monopolist, provided the relevant marginal revenue curves intersect the marginal cost curve in its vertical portion.

Figure 1 - Cost of Charging λ_H .

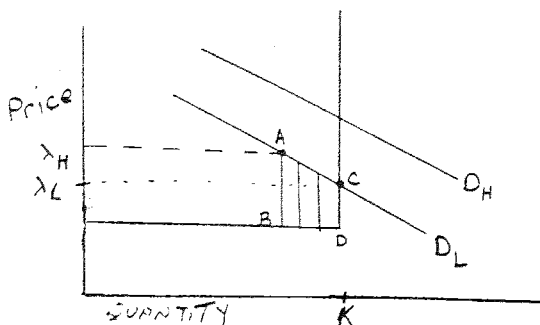


Figure 1 illustrates the cost of using price to allocate. There may be an alternative method of allocation that avoids or lowers this cost. We usually never have occasion to consider such an alternative because of the assumption of costless use of a price system.

To illustrate the point, let me use an extreme example. Suppose that the seller has perfect knowledge about the relative demands of his buyers. For example, suppose that the seller knows that he has only two buyers, each of whom are identical in their demands. Then, the firm knows exactly the efficient allocation even though it has no idea whether the market clearing price is λ_H or λ_L . The firm should give $\frac{K}{2}$ of its output to each buyer. By using its knowledge about its customers, the firm is able to achieve the efficient allocation of resources even though it has no idea of the correct market clearing price. It is precisely this simple point -- that knowledge about buyers assists efficient allocation -- that is ignored in most analyses but

provides the key to understanding the role of marketing departments, and the behavior of many firms and industries.

The simplification used in the preceding example of perfect rationing ability is obviously unrealistic. If it is costly to use the price system to allocate, it is likely to be costly to use other non-price methods and vice versa. But the simplification is a useful point to begin the analysis and is relaxed later. Notice that these simplifying assumptions -- costly use of a price system but costless use of non-price methods -- is the precise opposite of the assumptions usually made.

In the usual model, the price mechanism produces the optimal allocation and simultaneously generates enough revenue to finance the firm.¹² Once there is a cost to using the price system, there is no longer a unique correspondence between the achievement of the optimal allocation and the financing of the firm. For example, the firm could theoretically charge a zero price and achieve the efficient allocation in the simple model just presented. However such a firm would not have any incentive to remain in business.

One solution to the firm's problem of raising sufficient revenue in the context of the preceding example is to have the firm charge λ_L . Notice, from Figure 1, that if the

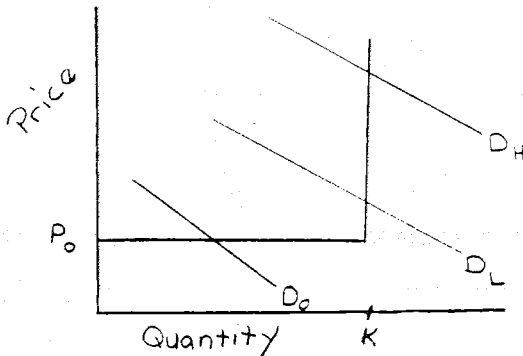
12. See Sheshinski and Dreze (1976) for an analysis of competition in the presence of fluctuating demand.

firm charges λ_L , then it does earn revenue in excess of variable cost and its use of the price system causes no potential for deadweight loss. As long as price is less than or equal to λ_L , no deadweight loss is created. But what if the revenue generated is not enough to cover the fixed costs F of the firm and what financial incentives are there for the firm to allocate efficiently?

In order to begin to address these questions, let us make the model of the firm a bit more realistic. The production technology of the firm will still be constant returns to scale up to some capacity K and require the expenditure of a fixed cost F . As before, there will be situations in which the firm must set a single price without knowing whether the demand curve is D_L or D_H . If demand is either D_L or D_H , we say that demand is in state 1, so that state 1 contains two possible outcomes (low and high) for demand. As before, we continue to make the simplifying assumption that the firm knows that its customers are identical so that the firm can, if it chooses to, achieve the efficient allocation. We also introduce another state of demand, state 0, knowable in advance to the firm so that the market clearing price can be set if the firm chooses to do so. The demand curve in state 0 is D_0 and is composed of the same buyers as in state 1. The configuration of the demand curves in state 0 and state 1 is illustrated in Figure 2. We assume that state 0 and state 1 occur with equal

probabilities and that within state 1 that D_H and D_L occur with equal probabilities.

Figure 2 - Demand Curves



We want to investigate how the firm should price in order to cover its fixed cost. We first analyze how a social planner would solve this problem. As usual, competition will force precisely this discipline on firms. The problem we seek to solve can be stated as follows: choose a method of pricing goods in state 0 and state 1 such that fixed costs are covered and deadweight loss is minimized. This is a variant of a standard problem in public finance. The new twist is that the price in state 1 is independent of whether the demand realization is D_H or D_L .

The optimal solution to this problem is to use lump sum taxes and marginal cost pricing or some other non-linear pricing scheme. As in the usual public finance context, this perfect theoretical solution is usually impractical and

is not common. For example, when buyers pay lump sum fees, there are incentives for buyers to merge and aggregate purchases to avoid the lump sum fee. Here, I want to examine a relatively simple and common form of pricing, the charging of unit prices that differ over demand states. The motivation for studying this form of pricing is that in my earlier work (Carlton (1986))¹³, I found it common that a buyer would face a price and then would choose quantity. I discuss non-linear pricing later on. Let us now solve the social planner's problem -- which should remind the reader of the usual Ramsey pricing problem.

The problem is to choose taxes, t_0 and t_1 , to add to marginal cost c in each state so as to cover cost F and create the minimum expected deadweight loss. Mathematically we have

$$\min_{t_0, t_1} \frac{-1}{2} \int_{Q_0(c)}^{Q_0(c+t_0)} [P_0(q) - c] dq - 1/4 \int_{K}^{Q_1^L(c+t_1)} [P_1^L(q) - c] dq$$

$$\text{s.t. } \frac{1}{2} t_0 Q_0(c+t_0) + 1/4 t_1 Q_1^L(c+t_1) + 1/4 t_1 K = F,$$

where

$Q_0(\)$ = demand curve in state 0. This occurs with probability 1/2,

$P_0(\)$ = inverse demand curve in state 0,

c = constant unit cost,

$Q_1^L(\)$ = demand curve in state 1 - low demand. This demand occurs with probability 1/2 in state 1 (which itself occurs with probability

13. See note 7, supra at 9.

1/2),

$P_1^L()$ = inverse demand curve in state 1 - low demand,

$Q_1^H()$ = demand curve in state 1 - high demand. This occurs with probability 1/2 in state 1 (which itself occurs with probability 1/2),

K = capacity,

t_0, t_1 = taxes, and

F = fixed costs.

The first term in the objective function is the deadweight loss that results in state 0 from a tax t_0 . The second term is the deadweight loss that results in state 1 - low demand from a tax t_1 . There is no term corresponding to the deadweight loss arising in state 1 - high demand state. That is because as long as the price is below λ_H , the assumption of perfect rationing guarantees that the firm allocates its K units efficiently.¹⁴ The constraint expresses the requirement that the overall expected revenues cover the fixed costs.

14. Under some circumstances, t_1 may be so high that the price in state 1 will exceed λ_H (see Figure 2), and there would be a third term in the objective function. We will in general obtain deadweight loss in the high state of demand in state 1 when we relax the perfect rationing assumption later on, and so here maintain the simpler assumption that the price in state 1 does not exceed λ_H .

It is straightforward to solve for the optimal tax rates. (See Appendix Theorem 1.) The optimal tax rates, τ_0 , (t_0/P_0) , and τ_1 , (t_1/P_1) , satisfy

$$(1) \quad \tau_0 \propto 1/\epsilon_0, \text{ and}$$

$$\tau_1 \propto (1+f) 1/\epsilon_1, \text{ where}$$

ϵ_i = elasticity of demand in demand state i ,¹⁵ and

$$f = \frac{K}{Q_1^L (c+t_i)}$$

Notice that f is the ratio of K to Q_1^L and therefore always exceeds 1. As t_1 increases, Q_1^L falls and hence f rises. Therefore, f becomes larger as the required revenue, F , and required taxes increase. Moreover, f increases the more "spread apart" are the demand curves in state 1, since f increases as Q_1^L falls.

If $\epsilon_0 = \epsilon_1$, then (1) simplifies to

$$(2) \quad \tau_1 = (1+f) \tau_0.$$

15. I have assumed that the elasticity of demand is the same for the curves Q_1^H and Q_1^L . It is straightforward to relax this assumption.

Eq (2) shows that in general it is optimal to cause a pricing distortion in both state 0 and state 1 with the price in state 1 exceeding that in state 0.

It is straightforward to show that if there are several states of demand, i , each of which contain pairs of demand curves, Q_i^H and Q_i^L , $i \geq 1$, among which the firm cannot distinguish, the optimal taxes, τ_i , satisfy

$$(3) \quad \tau_i = (1+f_i) \tau_0, \text{ where } f_i = \frac{K}{Q_i^L}$$

If the states of demand are such that Q_i^L increases with i , then taxes will tend to increase with i since λ_i^L will increase with i .¹⁶

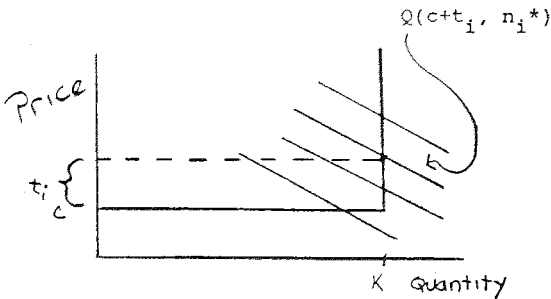
The preceding derivations produced very simple pricing rules. The reader may think that these rules come from the simple assumptions I have made about only two unobservable demand curves in state 1 and about the particular configuration of these curves as shown in Figure 2. It turns out that the simple pricing rules just derived hold under much more general assumptions -- as I now demonstrate.

Let each demand state i consist of a family of demand curves indexed by n , $\{Q_i(P, n)\}$. Let demand be increasing in n which has a probability density function $g_i(n)$. Demand is increasing in i in the sense that the probability density

16. It appears possible that the optimal taxes could decrease with i depending on the behavior of marginal revenue from taxation relative to the marginal deadweight loss. (See Appendix Theorem 2). This would lead to the possibility of countercyclical prices.

$g_i(n)$ has more weight in the upper tail as i increases. Let n_i^* be that n such that $Q_i(c+t_i, n_i^*) = K$. The interpretation of n_i^* is that it identifies all the demand curves ($n < n_i^*$) for which a deadweight loss occurs if price is elevated above $c+t_i$. For all demand curves for which $n > n_i^*$, there is no deadweight loss at $c+t_i$, since there is excess demand at that price and, by assumption, the firm can ration its fixed capacity efficiently. This choice of n_i^* is illustrated in Figure 3.

Figure 3 - Family of Demand Curves in State i



We are now ready to solve the more general problem. The social planner wishes to choose taxes, t_i , to minimize deadweight loss subject to a zero profit constraint. Mathematically the firm's problem is

$$(4) \min_{(t_i)} s_o \int_{Q_o(c,n)}^{\infty} g_o(n) \int_{Q_o(c,n)}^{Q_o(c+t_i, n)} \frac{Q_o(c+t_i, n)}{(P_o - c)} dq dn + \sum_i s_i \int_{n_i^*}^{\infty} g_i(n) \int_K^{Q_i(c+t_i, n)} [P_i - c] dq dn$$

subject to

$$s_0 \int_{-\infty}^{t_0} Q_0 g(n) dn + \sum_i s_i \left[\int_{-\infty}^{n_i^*} g_i(n) t_i Q_i(c+t_i, n) dn + \int_{n_i^*}^{\infty} g_i(n) dn t_i K \right] = F, \text{ and}$$

$$Q_i(c+t_i, n_i^*) = K, \text{ for all } i,$$

where

$g_i(n)$ = probability density of n in state i ,

$P_i(q, n)$ = inverse demand curve in state i , case n ,

$Q_i(P, n)$ = demand curve in state i , case n ,

s_i = probability of state i , and

all other terms were previously defined.

The first term in the objective function reflects the deadweight lost from charging $c+t_0$ rather than c in state 0.¹⁷ The second term represents the deadweight loss from charging $c+t_i$ instead of c in state i . Notice that deadweight loss occurs only in state i for $n < n_i^*$. The revenue constraints has three components. First is the revenue earned in state 0; second is revenue earned in state i when the amount demanded is less than K ; third is revenue earned in state i when the amount supplied is exactly K .

Solving (4) for the optimal tax rates, assuming the price elasticities of demand are equal across states, yields

17. For simplicity, I have assumed that in state 0 demand is always below K at price c .

$$(5) \tau_i = \tau_0 (1 + f_i) \quad \text{where}$$

$$f_i = \frac{K \int_{n_i^*}^{\infty} g_i(n) \, dn}{n_i^* \int_0^{\infty} g_i(n) Q_i(c + t_i, n) \, dn}$$

From the assumptions made, it follows immediately that $f_i > 0$. As noted earlier, it is expected that prices will vary procyclically, but it appears possible that price could actually move countercyclically depending on the rate at which deadweight loss and marginal revenue from taxation increase relative to each other as i and t_i increases. (See Appendix, Theorem 2).

The model I have just worked out is a relatively simple one but it contains three interesting implications.

1. The amplitude of price fluctuations tends to be smaller than that which would occur in the usual market clearing model, with price being higher in low demand states and lower in high demand states.
2. Even under the perfect rationing assumption, there is a deadweight loss that arises from using price to raise revenue to finance the firm. This deadweight loss rises the greater the "spread" of the family of unobservable demand curves in each observable state. Analogous to theories in public finance, it follows that the firm may use non-linear pricing schemes to finance its output.

Since these schemes are difficult to enforce, a reason for vertical integration arises, provided the vertically integrated firm is better able to enforce a non-linear pricing scheme among its divisions than it could if it were charging non-linear prices to independent agents.

3. The firm and customer benefit if there are more products for the firm to "tax." This follows immediately from the theory of public finance which teaches that the deadweight loss from raising a given amount of revenue declines as more products are taxed. Therefore, there is an economy of scope in the financing of the firm. This economy of scope is unrelated to both production and marketing considerations. A firm could achieve this economy of scope if it sold multiple products simultaneously or if it sold the same physical product at different points in the business cycle to the same customer. This economy of scope provides one reason for buyers and sellers to be "married" to each other for long periods of time. The seller commits to supply the buyer and the buyer commits to purchase from the seller over an extended time period. The implication is that a steady customer may well be treated differently than a sporadic one.

So far, we have examined the socially optimal behavior of a single firm. We now address the following questions:

- a) how is equilibrium determined?;
- b) what happens if rationing is not perfect?;
- c) how will the firm choose its customers when customers are not homogeneous?;
- d) how can a firm invest to improve its rationing ability?; and
- e) what determines the optimal K for a firm?.

IV. Equilibrium

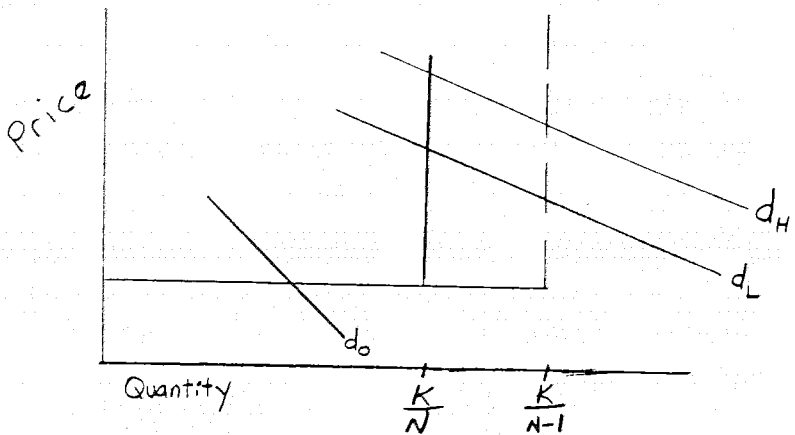
A. Competition

I now describe the competitive equilibrium for the simple model. Equilibrium requires buyers and sellers to contract with each other in advance for price and quantity. The price and quantity allocations depend on the observed state of demand. Competition forces firms to offer a price and ability to allocate that maximizes consumer surplus and forces firms to earn zero expected profits.¹⁸

18. The equilibrium is a long run concept in which entry erodes profits and all buyers and sellers are married. A more general model could have an equilibrium in which a steady stream of transient buyers appear, buy in a spot market, and leave. These transient buyers would bear the cost of using the price system and would pay on average a higher price than the married buyers. Without the ability of married buyers and sellers to
(Footnote Continued)

In order to focus attention on the key aspects of equilibrium, consider the demand curves, d_o , d_L , and d_H , of a customer of an individual firm and the supply available to serve that customer. For simplicity, we initially take all consumers to be identical. If a firm's capacity is K , then it is as if each of N customers of the firm have a mini-firm with capacity of K/N . As the number of customers served by a firm decreases, the capacity available to serve each customer increases. For example, the capacity devoted to serving each customer rises from K/N to $K/N-1$ as the number of customers decreases from N to $N-1$. See Figure 4.

Figure 4 - Per Capita Demand and Supply



(Footnote Continued)

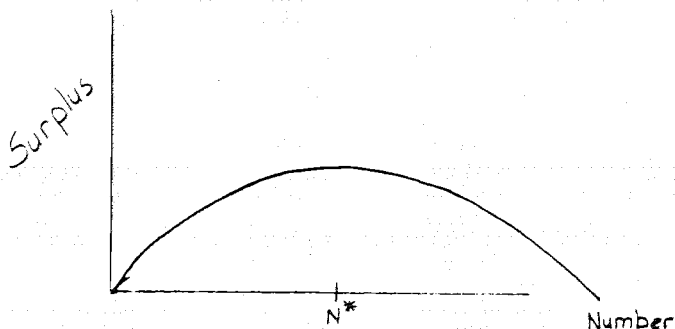
monitor each other, problems with a core not existing could arise. See Telser (1978).

A decrease in the number of customers at a firm has two opposing effects. First, it means that there are fewer occasions when rationing is required and during those times when customers are on allocation, more of each customer's demands can be satisfied. This raises consumer surplus. Second, because the same amount of revenue must be raised to cover the fixed costs regardless of the number of customers, the "tax burden" on the remaining customers increases and reduces their consumer surplus. For some N^* , the firm, constrained to earn zero profits, maximizes consumer surplus per customer. See Figure 5. Any firm with either fewer or more customers than N^* would be unable to successfully compete for customers. What is interesting about the equilibrium is that the size of a firm matters to its customers because it affects the firm's ability to finance itself and efficiently allocate resources. Shortly, we develop a theory of customer mix, once the assumption of identical consumers is relaxed. Once N^* is determined, eq. (3) determines the optimal pricing in equilibrium.

One interesting comparative static exercise is to compare the equilibrium to the one that results when use of the price system is not costly. I now explain why the optimal customer size tends to be smaller when use of the price

system is costless.¹⁹ (See Appendix Theorem 3 for more detailed discussion). Consider the consumer surplus curve in Figure 5.

Figure 5 - Per Capita Consumer Surplus



The curve initially rises because as the number of customers increases the gain in utility from a reduced per capita "tax" burden exceeds the increase in deadweight loss from having less output per customer to allocate. Notice that the allocative efficiency properties of a system with perfect rationing could be identical to the (costless) price system but for the financing constraint. But the financing constraint imposes a deadweight loss in each state of nature because the tax causes a marginal distortion. This marginal

19. Although the analysis assumes a fixed K , it is possible to reinterpret the results as applying to a situation where each unit of capacity costs F and K is freely chosen. All theorems would then apply to K/N .

distortion is not present when the price system is costless and a different price can be charged for every realized demand state. Deadweight loss falls as the number of states in which to charge taxes increases and the marginal distortions fall on average. Therefore, as more customers are added, the gain from being able to lower "taxes" saves more deadweight loss when firms are using rationing (since existing average distortions are larger). This creates an incentive for these firms to be larger than they would be if using the price system were costless. This is especially true the wider the divergence between the unobservable demand curves within any one observable state of demand because that is a situation with large marginal distortions. Therefore, the number of customers served by a firm tends to decrease as the cost of using the price system falls, and there will be more firms. In this sense, a costless price system and atomistic competition go hand in hand.

B. Monopoly

The monopoly solution is similar to the one just derived. A monopolist maximizes his profit by adopting a pricing scheme and quantity allocation that efficiently allocates the goods he does sell. This implies that the monopolist's prices, though generally higher than a

competitor's, will also satisfy equations like (3).²⁰ A monopolist charges higher prices than competitors, and, as a result, he diminishes his need to rely upon rationing. The fact that the monopolist can charge higher prices than would prevail under competition means that f in (3) is higher under monopoly and, therefore, that there would tend to be more procyclical price movement for a monopolist.

V. Relaxing the Assumption of Perfect Rationing

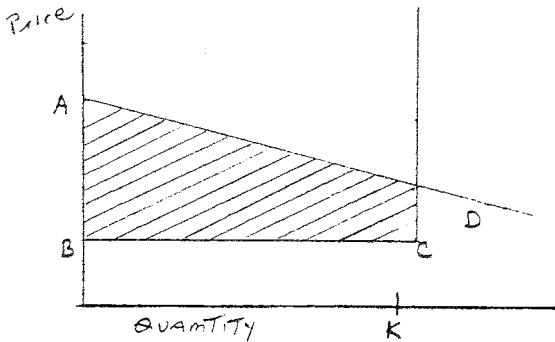
The assumption that firms have the ability to perfectly rank their customers' relative need for a good is obviously unrealistic. In reality, a firm will have difficulty allocating its output capacity, K , among its customers and will not, in general, be able to duplicate the efficient allocation that would result if use of the price system were costless.

When non-price methods entail costs, then the problem facing the firm becomes more complex. Let R stand for the firm's rationing ability. It represents that fraction of maximum consumer plus producer surplus that a firm can achieve for any given demand curve. For example, in Figure

20. The monopolist maximize profits subject to a utility constraint. If maximum profits are Π^* , then the optimal solution will maximize utility subject to the constraint that profits equal Π^* . This maximization problem is similar to the one solved by (3).

6, R is the fraction of $ABCD$ that the firm can generate. So far, I have taken R to equal 1.

Figure 6 - Consumer Plus Producer Surplus



There will be many factors that influence R . For example, R will depend upon the number and heterogeneity of customers, the knowledge that a firm has about its customers and the price a firm charges. The more numerous and heterogeneous are customers, the more complicated it is for a firm to know which customers have the greatest need for its product. The more knowledge a firm has about its customers, the easier it is for the firm to allocate efficiently. One function of salesmen is to understand the needs of their customers, and in times of shortage, make sure that those customers who need the good the most obtain it.

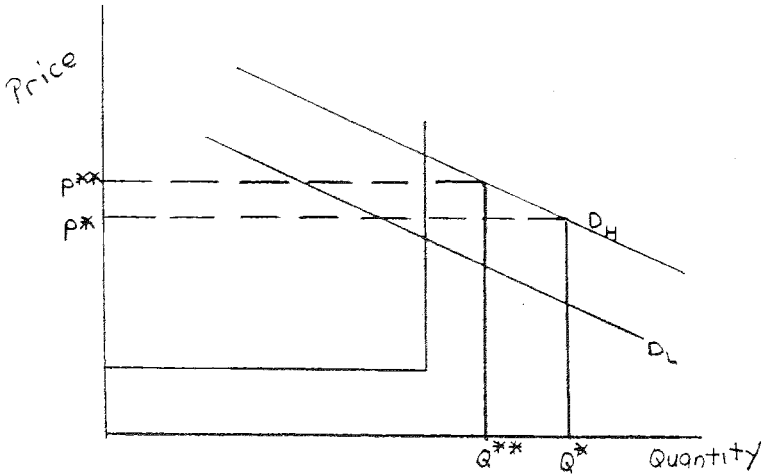
When it is costly to allocate by non-price methods, the efficiency of the non-price method of allocation will depend

on the price charged. By reducing the amount of excess demand in periods of shortages, a high price reduces the number of customer (or units demanded) over which non-price allocation methods are used. It is as if price is used as the first screen to identify the most likely candidates to receive the good, and then non-price methods are used for the final allocation. As price rises, the deadweight loss from inefficiently reducing consumption in low demand states (i.e. the cost of using the price system identified in Figure 1) increases, but the cost of inefficiently allocating by non-price methods the excess demand in high demand states falls. The general point is that it is efficient to use both price and non-price allocations when each are costly to use.

To see this point, consider the following example illustrated in Figure 7. Suppose that one state of demand has two unobservable states. If a price P^* is charged and D_H occurs, Q^* units are demanded and the firm must decide which K of those to satisfy. If instead a higher price P^{**} is charged and D_H occurs, only Q^{**} is demanded. The firm has a better chance of generating the maximum consumer plus producer surplus in the latter case since the lower value users have been screened out by the higher price. Notice that although the higher price P^{**} improves the efficiency of the non-price rationing if D_H occurs, the higher price creates a larger deadweight loss if D_L occurs because the

higher price drives consumption farther below K . This identifies the trade-off between the use of price and non-price methods of allocation. In general, it will be optimal to use both price and non-price methods to achieve efficient allocation when the use of each creates some deadweight loss.

Figure 7 - Interaction Between Prices and Non-Price Rationing When Both Are Costly



As a firm's rationing ability improves, it tends to rely less on price to achieve efficiency and its pricing tends to become less procyclical. To see this point intuitively, notice that if a high price is not needed for efficiency reasons to help allocate goods in times of shortage, then it pays to reduce t_1 and shift taxes to the other (lower) observable states of demand. More rigorously, the firm faces the following problem:

$$\begin{aligned}
 & Q_0 (c+t_0) && Q_1^L (c+t_1) \\
 \min_{t_0, t_1} & \frac{-1}{2} \int_{Q_0(c)}^{P_0(q)-c} [P_0(q)-c] dq - \frac{1}{4} \int_K^{P_1^L(q)-c} [P_1^L(q)-c] dq \\
 (6) & -\frac{1}{4} (1-R) \int_K^O [P_1^H(q)-c] dq,
 \end{aligned}$$

$$\text{S.t. } \frac{1}{2} t_0 Q_0 (c+t_0) + \frac{1}{4} t_1 Q_1^L (c+t_1) + \frac{1}{4} t_1 K = F,$$

where all terms have been previously defined.

The only difference between this problem and the one solved earlier is the presence of $(1-R)$ multiplying the last term in the objective function. When $R=1$, (6) is identical to the earlier problem.

Solving (6) yields the optimal solutions for (t_0, t_1) .

It is possible to prove that under certain assumptions

$\frac{\partial t_0}{\partial R} > 0$ and $\frac{\partial t_1}{\partial R} < 0$. (See Appendix Theorem 4.) This means

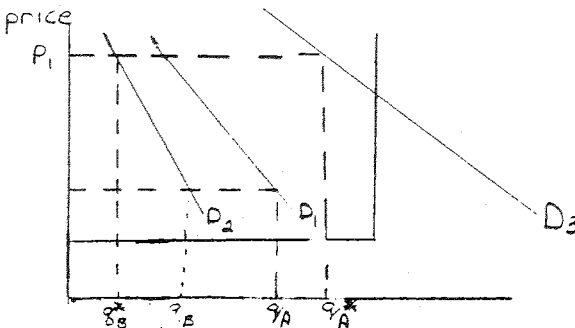
that the firm relies less on price to clear markets in peak demand periods as the firm's rationing ability improves. It also implies that prices becomes less procyclical as rationing ability improves.

Since customers are no longer assumed to be identical and since the firm has some knowledge about each customer, it follows that the optimal price and allocation policy varies from customer to customer. In an effort to obtain an efficient allocation, the firm prices to screen out low value demanders before it relies on non-price allocation methods. If a firm has different knowledge about each

consumer then the firm should charge each consumer a different price.

To understand this point, suppose that the firm is in a period of shortage and that it has two customers. See Figure 8.

Figure 8 - Charging Consumers Different Prices



Customer A has a demand curve that is either D_1 or D_3 , while Customer B has a demand curve that is D_2 . The optimal allocation in the case where the demand curve of Customer A is D_1 , is q_A to Customer A and q_B to Customer B, where $q_A + q_B = K$. In the case where the demand curve of Customer A is D_3 , the optimal allocation is q_A^* to Customer A and q_B^* to Customer B, where $q_A^* + q_B^* = K$. If the same price were charged to each customer and that price were below P_1 , then the rationing efficiency of the firm is seriously impaired when Customer A's demand is high and equals D_3 . The reason is that so many units are demanded by Customer A that it makes it difficult to allocate the goods amongst customers when rationing is imperfect. In fact, it is essential to

optimal operations that different customers pay different prices when both rationing and the price system are costly to use. Different customers impose different costs on the firm's allocation ability and, as should be expected, different costs should give rise to different prices. The stochastic buying characteristics of a customer affect the cost of servicing that customer. A steady customer imposes less costs on a firm than one who has variable demands and therefore should receive a lower price. It is a consequence of the relaxation of costless operation of the price system that requires different customers pay different prices in order to assure efficient operation. Price differences among customers are a natural consequence of market operation when organized auction markets don't exist.

VI. The Role of the Firm in Facilitating Trade Between Different Customers

This section explores in more detail the effect of differences among consumers. The main point is to show how a firm acts to facilitate trade amongst customers when there is a costly price system and when the firm can identify different groups of consumers. This function of the firm has received little attention in the current treatments of the theory of the firm. It is a role that could reasonably be identified with the marketing departments of large firms. I first allow customers to differ in the elasticity of

demand in state 0. I then introduce uncertainty in state 1 and analyze the implications of having a customer whose demands are only partially correlated with the rest of the population.

A) Different Elasticities

Suppose that there are two groups of customers, Group A and Group B, who differ in only one respect; they have different elasticities of demand in state 0. Their demands in state 1 are perfectly correlated. To fix ideas, suppose that one group, Group A, has a perfectly inelastic demand in state 0. Suppose that a firm just has consumers from Group A as its customers. From the previous analysis, it follows that the optimal solution is for the firm to charge a price λ in state 1 such that $Q_1^L(\lambda)=K$ and charge a large enough fee in state 0 to cover the fixed costs. Moreover, these consumers will have the same consumption in the low state as in the high state of state 1. (Indeed, this solution is analogous to one involving a two-part tariff.)

Now consider a firm that caters solely to Group B. These customers have a finite demand elasticity in state 0, so that, as in the previously worked out models, they wind up paying a price in state 0 in excess of constant marginal cost, c , and their consumption in the low state in state 1 is below that in the high state in state 1. Compared to firms catering to Group A, firms catering to Group B will

have less capacity per customer (K/N smaller) because they are less efficient at raising money to pay for capacity.

The question we now address is whether a firm should serve both customer groups on the assumption that the firm can separately identify each group. That is, is there an economy of scope in serving both groups? The answer is generally yes. To see why, consider state 1. We know that in the case of a firm serving only Group B there is excess capacity in the low state of state 1. This excess capacity arises because one consequence of a costly price system is that the price in the low state of state 1 chokes off too much demand.

Now consider a firm catering solely to Group A customers. These Group A customers could well want to consume more at a marginal cost of c than the amount that they are currently consuming in that state. Moreover, the shadow price of the Group B consumers in the high state of state 1 is higher than that of a Group A consumer (because K/N is higher for the firm specializing in Group A consumers). Hence, there are gains for trade if a mechanism could be devised. The obvious problem is that if the price system is costly, it is hard for the two groups of customers to trade in the two states of state 1 that by assumption can't have different prices because of the cost of creating markets. It is precisely in this situation that the firm can step in and facilitate trade between the two groups.

Recall that although a firm cannot charge different prices to a customer in the two unobservable states of state 1, the firm can use its knowledge of customers' needs when rationing is needed. The optimal arrangement is for the firm to serve the two groups, to use some of the available capacity of a Group B customer in the low state of state 1 to serve a Group A customer, and to switch the consumption of a Group A customer to a Group B customer in the high state of state 1. But how can the firm find it in its financial interest to do this? After all, it can't vary price between the two states in state 1. The answer is that the firm can now raise (or lower) the (lump sum) tariff in state 0 to Group A customers. These Group A customers can be thought of as "buying" consumption in the low state of state 1 and paying for it by "selling" consumption in the high state of state 1 and by paying (or receiving) money in state 0. The point is that the firm serves the function of allowing consumers to transact across states when the price system is too costly to be used. (The exact terms of how Group A and Group B fare will depend on their relative population sizes.) Therefore, in general, we expect to see firms serving multiple consumer groups with each group paying different prices and being rationed differently. As before, the firm's rationing ability, R , will determine how efficient the reallocation can be. However, even if $R=1$, the firm will not, in equilibrium, reproduce the allocation of a

costless price system and it will still typically be the case that the price in state 0, P_0 , will exceed c for the Group B consumers.

B) Differing Stochastic Demands

Suppose now that customers differ in the demands that they have in state 1. For example, each customer i could have a random term that equals $\rho_i \epsilon + \delta_i$ where ϵ is a common random component, δ_i is an independent and identically distributed random variable, and ρ_i is a correlation coefficient ($0 < |\rho_i| < 1$) for consumer i .

Suppose first that $\rho = 0$ for all customers. By putting together a diversified portfolio of all customers, the law of large numbers guarantees that the likelihood of rationing in state 1 will go to zero. In this setting, P_0 approaches c and P_1 approaches $c + F^*$, where F^* is the fixed cost of capacity per customer which equals the average amount demanded in state 1. Each firm has a completely diversified set of consumers. There is no benefit to a firm from being skilled at rationing, just from being diversified.

Now suppose $\rho_i \neq 0$. To fix ideas, suppose that there are two groups, Groups A and B with ρ equal to ρ_A and ρ_B respectively where $\rho_A > \rho_B > 0$. Suppose ϵ takes on just two values with equal probability, $-a$ and a . Then, the demand curve of Group A lies above that of Group B when $\epsilon > 0$ and below that of Group B when $\epsilon < 0$. Just as before, there may be gains for

trade between the two groups. The firm will act to shift more goods to Group A in the high state of state 1 and to Group B in the low state of state 1. The firm accomplishes this reallocation by using its knowledge of customers and its rationing ability. Each customer pays a price in state 0 and faces a rationing rule and price in state 1. The prices and rationing rule vary by ρ_i . Notice that as long as $\Sigma \rho_i \neq 0$ for the group of customers, then there will be aggregate uncertainty in the demand curve facing the firm in state 1 and the need for rationing will persist even when each firm is diversified across customer groups.

We have already discussed how R will be affected by the number and heterogeneity of customers. As long as the need for rationing persists, diversification is costly because it affects R through the number of customers. This leads to a finite firm size even in the presence of constant returns in the building of capacity. As a customer's ρ increases, the firm's ability to ration declines (because the customer adds variance to demand, making more rationing necessary in the high state of state 1). We therefore should expect that customers are charged for ρ so that their average price paid rises with ρ . In summary, a customer's stochastic buying characteristics influences the price he pays.

To illustrate these ideas, suppose Group A has $\rho_A = 0$ and $\delta_A = 0$. Such a group can use the price system efficiently because there are no unobservable states in state 1.

Suppose Group B has $\rho_B > 0$. That group has to be rationed sometimes for reasons already discussed. Half the time Group B values the good more than Group A and vice versa in state 1. Group A would willingly join a firm that uses rationing even though Group A's demand curve is unchanged between the high and low state in state 1. The reason is that Group B can "trade" with Group A if they are in the same firm. The firm would charge Group A a price below $c+F^*$ which allows them to expand their consumption in the low state of state 1 but ration them in the high state of state 1. This arrangement enables Group B to expand consumption in the high state of state 1. Again, the firm is facilitating trade between two groups in a situation where the trade could not otherwise occur because of the non-existence of a separate price in each of the two states in state 1.

This model emphasizes that in state 1, it is the firm's rationing ability combined with the magnitude of the common shock that determines the unavoidable inefficiency that results from having a costly price system. Conditional on receiving a large positive shock, ϵ , the system does poorly in that this realization of ϵ is one that creates lots of deadweight loss ex post (though not ex ante in expected value). In response to large increases in demand, prices don't rise and rationing does a poor job at allocating goods. If such shocks were common there would be more use of prices. Therefore, the consequence of having a costly

price system is that the allocation mechanism is not very good at coping with large unexpected positive shocks.²¹

VII. Empirical Evidence

In this section, I report on some available empirical evidence that supports many of the key assumptions and implications of the theory. The main assumption and implications of the theory are as follows:

1. It is costly to use the price system for some goods so that rationing is optimal when demand is high. Customers will be charged different prices depending on the stochastic nature of their demand. For example, customers who buy only at peak times may not be able to obtain the good with the same reliability or at the same price as steady customers.
2. The amplitude of price fluctuations tends to be smaller than that which would occur in the usual market clearing model, with price being higher in low demand states and lower in high demand states.
3. The system of market operation involving rationing is not sustainable during extended periods of

21. Negative shocks do not raise the same problems because rationing replaces the price system in state 1 only for positive shocks.

chronic excess capacity. In such a case, price will fall to marginal cost because customers see no benefit to paying higher prices to guarantee supply in the future.

It is difficult to obtain systematic evidence on the importance of non-price rationing from published data sources because it would be impossible to examine data on price and quantity and infer rationing. Instead we must rely on evidence more akin to industry studies of particular industry and accounts in the business press.

There have been several press accounts documenting the imposition of allocations, quotas, and non-price rationing. By non-price rationing, I mean instances where customers are put on allocation and are not able to obtain the full amount of their demand. For example, the Wall Street Journal reported (7-7-87) that plastics producers (polypropylene) were accepting no new customers and had placed quotas on their deliveries to existing customers. Other industries where non-price rationing has been reported include plastics, computer chips, chemicals and various metals.²²

22. For some newspaper accounts indicating the existence of non-price rationing, see Iron Age, 1/6/75, p. 90 (non-ferrous metals), Wall Street Journal, 4/17/84, p. 1 (aluminum, electronic parts, metal fasteners, and gypsum board), Wall Street Journal, 4/17/86, p. 1 (titanium dioxide), N.Y. Times, 1/30/84, p. D-1 (microprocessor chips), Wall Street Journal, 6/3/83, p. 23 (personal computers), N.Y. Times, 6/29/83 (semi-conductors), Wall Street Journal, 10/23/86, p. 1 (compact disks), Wall Street Journal, 7/7/87

(Footnote Continued)

There have been several studies documenting a two-tiered pricing system in many metal markets.²³ For numerous metals, (e.g. aluminum, cobalt, copper, nickel, molybdenum) it was typical that there were two prices - one for spot buyers and one for steady buyers. The price that the steady buyer paid (often called the producer price) was typically different than the spot price. Consistent with the theory, the spot price was much more volatile than the price to steady buyers who comprised the bulk of the market.

One of the rare reports analyzing these types of issues found: ".....producer prices have not as a rule risen sufficiently to clear the market during periods of scarcity, and producers have generally limited their price cuts in times of excess capacity. Because producers reward customer loyalty, those who purchase from a producer at its posted prices usually have access to product supplies in times of shortage, whereas those who rely on spot purchases risk continued dependence on the spot market at those times."²⁴ The report also found, consistent with the theory, that the two-tiered system of pricing tends to disappear when excess

(Footnote Continued)

(plastics), Chemical and Engineering News, 4/27/87, p. 17 (petrochemicals).

23. See e.g., McNicol, (1974), Peck (1967), Burrows (1972), Charles River Associates (1967, 1970) and U.S. Senate Hearings (1956).

24. P.4, Charles River Associates (1986).

capacity is expected to persist. For example, in nickel, copper, and aluminum, there has been excess capacity for several years. In those industries, the two-tiered pricing system has vanished.

The theory predicts that one determinant of a customer's price is the correlation of that customer's demand with that of the firm. Schwieterman (1990)²⁵ has tested this theory for airline pricing.²⁶ His findings are striking. There is a strong relationship between fares and stochastic demand patterns. For example, someone travelling on a restricted fare (Max Saver Fare) has a correlation with the firm's demand of about one third that of other passengers travelling on the same flight. The lower correlation is consistent with the passenger on a Max Saver Fare receiving a substantial discount.²⁷

-
25. Schwieterman, J., The Consequences of Airline Price Deregulation: A Hedonic Price Approach, Ph.D. Thesis, University of Chicago Public Policy, 1990.
 26. This airline case is a bit different from the model since there is no long run relationship necessarily between fliers and airlines (though frequent flier programs attempt to create such a relation). However, as long as tickets are purchased ex ante, the theory of the model implies that a buyer's stochastic buying characteristics should influence the price.
 27. I recently had the opportunity to use confidential data to test the theory for a manufacturing industry by running econometric regressions of the form:

$$\text{Price} = f(\text{product characteristics, volume of buyer order, volume of one year of buyer's orders, buyer's correlation with firm's demand } (\rho))$$

(Footnote Continued)

VIII. Implications for Firm Behavior

The models of costly use of the price system and costly use of non-price allocations can be viewed as defining the implicit technologies for allocation by price and non-price methods. Resources are used (i.e. costs imposed) when a price system is used and when rationing is used. Because we know the technology used to allocate, we are able to understand the incentives to improve these technologies. I believe these incentives lead to an improved understanding of several aspects of the modern corporation, especially the marketing function.

A) Optimal Customer Mix

We have already seen in the previous sections how customers who place different demands on a firm's rationing ability are treated differently and have examined the incentive of a firm to put together a portfolio of different customers. The theory here is akin to the theory of clubs

(Footnote Continued)

The econometric results relied on several thousand monthly observations across several hundred customers for a several year period. The coefficient on ρ was statistically significant and positive. The magnitude of the coefficient indicated that different stochastic buying patterns of customers caused price paid to vary by as much as 10 to 15 percent.

in which the size and composition of the club affect each member's welfare. To understand the advantage of a desirable customer mix, suppose there are several customers whose demands are random but are uncorrelated so that the sum of their demand curves is non-stochastic. This enables the firm to avoid the problem of being unable to know which demand state prevails since by construction the total demand curve is always same.

In general, when a firm is considering adding a new customer, the firm calculates the additional cost that the new customer would impose on the firm's ability to ration to others. That cost tends to be greater the more positively correlated are the demands of the new customer with existing customers. The gain to the firm is the additional revenue (and hence reduced revenue burden on existing customers) that the new customer provides. It is the covariance of the new customer's demand with aggregate firm demand that is the key variable that determines how price is set to an individual customer.²⁸ This theory of customer mix is analogous to optimal portfolio theory. It implies that "goods" are much more heterogeneous than they may initially appear. A pound of product purchased steadily is a different product than a pound purchased irregularly. It also implies that one goal

28. It is more complicated if the elasticity of demand differs among customers. Moreover, the variance of demand could matter if buyers are large.

of the marketing department of a firm is to put together a desirable customer mix.²⁹

B) Improving the Firm's Ability to Allocate

The knowledge that a seller has of its buyers plays a crucial role in determining R. The longer a seller knows its buyers, presumably the better the ability of the firm to use non-price rationing. By assigning each salesman fewer customers or by sponsoring occasions wherein buyers explain their needs to sellers, the firm can improve its ability to use non-price rationing.

The firm can improve its ability to use the price system by improving its ability to distinguish among different aggregate demand states. By improving its overall sales and marketing force, the firm can become increasingly refined in its ability to discern different demand states and can thereby charge different prices for the different demand states. By investing to refine its ability to distinguish between various states of demand, the firm reaps the benefits of being able to use price to allocate goods at low

29. Buyers could band together into desirable groupings or be banded together by a third party (e.g. brokers, middlemen) to achieve the same desirable result of stabilizing demand. For example, advertising agencies often package together different demanders for TV advertising to produce a stable demand and then purchase for these demanders blocks of advertising time from TV networks.

cost. That is, the firm invests in order to lower its cost of using the price system.

The incentive to invest in order to use the price system more efficiently and to use non-price rationing more efficiently explains the importance of the marketing and sales function in modern corporations. It is impossible to tell whether a firm's investments in marketing are more productive in improving its rationing ability or in lowering its costs of using the price system. A firm may rely more or less on price as its marketing knowledge improves. One prediction is that a firm with many transient customers may do better improving its ability to use price because it may be easier to learn about aggregate demand than about customers' relative demands. In contrast, a firm with a few steady customers may find it easier to learn about relative demands and improve its non-price rationing ability. The firms in my previous study (Carlton (1986))³⁰ that I cited as evidence in support of non-price rationing were firms who sold intermediate manufactured products to large steady customers.

C) Economies of Scope in Marketing

A seller's knowledge of a customer is an important influence in determining R. If the firm knows the relative demands of a customer for one product, it may well have

30. See note 7, supra at 9.

insight into his relative demands for related products. For example, a lawn store has a high demand for both grass seed and fertilizer during peak demand times. Because of this, there is a natural economy of scope in marketing, and it therefore makes sense for a single firm to be in charge of the non-price allocation for several products when the products are complementary. This economy of scope is likely to be most important when inputs are used primarily for one industry so that the demands for the various inputs are highly correlated. The knowledge spillover also creates an economy of scope when the price system is used. As long as demands for different products are correlated, the firm selling both products can do a better job of using the price system than a firm selling one product because the firm selling both products is able to improve its overall knowledge about the state of aggregate demand about one product by observing the demand of the other. For example, marketing personnel may be able to pool information across products to better discern the current demand state of one product.

D) Hostility to Futures Markets

The knowledge that a firm possesses to use both price and non-price methods to allocate goods is a valuable asset and one on which the firm earns a return. If an alternative allocation mechanism were to appear, the firm's valuable

asset would decline in value. Firms and organized markets are competitors in producing "allocations" and like any competitor, the firm does not like competition. This phenomenon explains the hostility firms often have toward the introduction of a new organized futures market in their industry.³¹

E) Capital Investment

The marketing knowledge that the firm possesses to distinguish different states of demand and to determine the relative demands of customers will affect the firm's decision of how much capital to invest in. I expect that firms with higher values for R will have a higher ratio of capital to customers, will earn higher profits, and will have more customers than firms with lower values of R . This suggests that new firms that will tend to have low values of R will have an optimal scale below that of an established firm with a high value of R . Moreover, in contrast to established firms, entrants will have to rely more on price fluctuations to allocate goods despite the large deadweight loss that such fluctuations may entail. Entrants may also rely on more flexible production technologies even if they are more

31. For an example of hostility to a new futures market, use the discussion of the reaction of aluminum firms to the establishment of the aluminum futures market in American Metal Market, January 6, 1978, p.9.

costly than those used by established firms who have a steadier customer demand.

F) Swaps

In several industries, firms often swap product on an equal tonnage basis. For example, a paper firm with a customer in New York but a plant in California may swap 10 tons of paper with a firm with a plant in New York but a customer in California. These swaps often do not involve prices but do require equal tonnages. In fact, personnel in charge of arranging swaps may have little or no price making authority. There has been many reasons given for these swaps. (See Williamson (1983) and Joyce (1983))³². One reason for swaps that emerges from the theory of this paper is that swaps avoid requiring either firm to incur the costs of ascertaining the market price. This reason for swaps makes sense only if prices are highly correlated in the areas involving swaps and if it is not easy to determine the market price in each area. This theory predicts that swapping and non-price allocations are related phenomena.

In summary, recognition of the cost of using both the price system and non-price allocations helps explain the

32. Williamson, O., "Credible Commitments: Using Hostages to Support Exchange," American Economic Review, 73, (September 1983) 519-540, and Joyce, J., "Why Firms Rely on Barter," Disc. Paper 83-3, Department of Justice (1983).

purpose of marketing in the modern corporation. Marketing consists of investments designed to improve a firm's knowledge about the level of aggregate demand and the relative demands of customers. This improved knowledge lowers the costs of using price and non-price methods of allocation. Economics of scope in knowledge acquisition explain the incentive for the firm to sell related products. The marketing ability of the firm will influence its investment decisions and its need to rely on price fluctuations.

IX. Conclusion

Relaxation of the assumption of costless use of a price system leads to a theory with several implications for optimal allocation, market structure and industrial behavior. The theory implies that rationing is to be expected and that a customer's stochastic pattern of demand will influence his price. Marketing departments add value by putting together a desirable portfolio of customers and by learning about their customers' needs and using that knowledge to allocate goods. The theory implies that one reason for a firm's existence is to facilitate trade amongst its customers. This theory explains some puzzling empirical evidence and leads to an improved understanding of industrial structure and the marketing functions of a firm.

BIBLIOGRAPHY

- Burrows, J.C., Cobalt: An Industry Analysis, (Lexington, Mass: D.C. Heath, 1972).
- Carlton, D.W., "Futures Markets: Their Purpose, Their History, Their Growth, Their Successes & Failures", Journal of Futures Markets, (Fall 1984) 4(3):237-271.
- Carlton, D.W., "The Rigidity of Price," American Economic Review, (September 1986).
- Carlton, D.W., "The Theory and Facts of How Markets Clear: Is Industrial Organization Useful for Understanding Macroeconomics," Chapter 15 in R. Schmalensee and R. Willig ed., Handbook of Industrial Organization, (1989). A revised version appears in Chapter 21 of D. Carlton and J. Perloff, Modern Industrial Organization, Scott-Foresmen (1990).
- Chandler, A., The Visible Hand, Harvard University Press, Cambridge, Mass (1977).
- Charles River Associates, Economic Analysis of the Copper Industry, Report submitted to the Property Management and Disposal Service, General Services Administration, (March 1970).
- Charles River Associates, Economic Analysis of the Molybdenum Industry, Report submitted to the Property Management & Disposal Service, General Service Administration (1967).
- Charles River Associates, Natural Gas Procurement Experience with Spot vs. Contract Pricing in Analogous Commodity Markets, 12/86.
- Coase, R., "The Theory of the Firm," Economica, 1937.
- Dixit, A., & Stiglitz, J., "Monopolistic Competition and Optimum Product Diversity," Bell Journal, (1977), 297-308.
- Joyce, J., "Why Firms Rely on Barter," Disc. Paper 83-3, Department of Justice, 1983.
- McNicol, D.L., "The Two-Price Systems in the Copper Industry" Bell Journal of Economics, 6(1) Spring (1975):50-73.
- Peck, M.J., Competition in the Aluminum Industry 1945-1958, (Cambridge: Harvard Univ. Press 1967).

- Schwieterman, J., The Consequences of Airline Price Deregulation: A Hedonic Price Approach, Ph.D. Thesis, University of Chicago Public Policy, 1990.
- Sheshinski, E., & Dreze, J., "Demand Fluctuations, Capacity Utilization & Costs," American Economic Review, 66(5), December (1976):731-742.
- Spence, M, "Product Differentiation & Welfare", American Economic Review, May 1976 407-414.
- Telser, L., The Theory of the Core, University of Chicago Press, Chicago, Il:(1978).
- U.S. Senate Hearings, Supply & Distribution of Nickel, U.S. Senate Hearings before Subcommittee on Small Business, 84th Congress 2nd Session (Washington: U.S. Government Printing Office, 1956).
- Weitzman, M., "Prices vs. Quantities," Review of Economic Studies, (October 1974).
- Williamson, Oliver, "Credible Commitments: Using Hostages to Support Exchange", American Economic Review, 73, September 1983, 519-540.

Appendix

Theorem 1: The optimal taxes are given in eqs. (1) and (3).

Proof: Minimizing the objective function following Figure 2 subject to the financing constraint yields:

$$Q'_0 t_0 = \lambda [t_0 Q'_0 + Q_0], \quad \text{or}$$

$$(1-\lambda) t_0 Q'_0 = \lambda Q_0 \quad \text{or}$$

$$\frac{t_0}{P_0} \frac{P_0 Q'_0}{Q_0} = \frac{\lambda}{1-\lambda} \quad \text{or}$$

(A-1) $\tau_0 \varepsilon_0 = \frac{\lambda}{1-\lambda}$, where λ is a negative LaGrange multiplier and a " ' " denotes differentiation with respect to price. Also,

$$Q^L_1 t_1 = \lambda [t_1 Q^L_1 + Q^L_1 + K], \quad \text{or}$$

$$(1-\lambda) t_1 Q^L_1 = \lambda [Q^L_1 + K], \quad \text{or}$$

$$\frac{t_1}{P_1} \frac{P_1 Q^L_1}{Q^L_1} = \frac{\lambda}{1-\lambda} \frac{Q^L_1 + K}{Q^L_1}, \quad \text{or}$$

$$(A-2) \quad \tau_1 \varepsilon_1 = \frac{\lambda}{1-\lambda} \left[1 + \frac{K}{Q^L_1} \right] \quad \text{QED.}$$

It follows from (A-1) and (A-2) that $\tau_1 = (1+f)\tau_0$ where $f = \frac{K}{Q^L_1}$, and that $\tau_1 > \tau_0$.

Q^L_1

Theorem 2: If there are several states of demand, i , and if Q_i^L increases with i , then t_i could increase or decrease with i .

Proof: The first order condition requires that

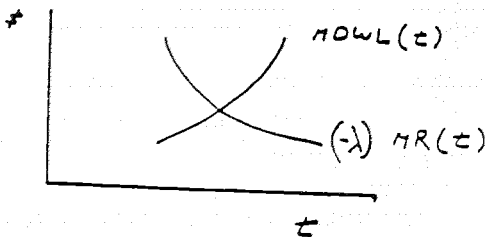
$$(A-3) \quad -t_i Q_i^{L'} = -\lambda [t_i Q_i^{L'} + Q_i^L + K]$$

The LHS is the marginal deadweight loss (MDWL) from increasing taxes, while the right hand side is the marginal revenue (MR) from increasing taxes times $(-\lambda)$. It is reasonable to assume

that

MDWL is upward sloping in t and MR is downward sloping in t , as illustrated in Figure A-1.

Figure A-1



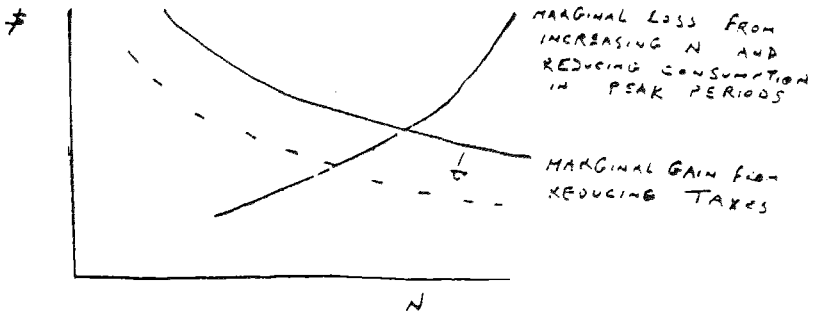
As i increases, the MDWL (t) curve could shift up or down, as could the MR (t) curve. Therefore, without further assumptions, it is possible for t_i to either increase or decrease with i .

It is possible to identify conditions under which taxes will move procyclically. For example, if demands are linear, then it immediately follows that t_i increases with i ($(-\lambda) MR(t)$ curve shifts up; MDWL (t) curve remains unchanged).

Theorem 3: The equilibrium size of a firm tends to be larger with rationing as compared to the case of costless use of a price system.

Proof: The equilibrium N occurs where average consumer surplus is maximized. There are two effects that determine the optimal N . First as N increases, the capacity available to satisfy the needs of a single customer during peak demand falls. This represents a cost. Moreover, because demand curves slope downward, the marginal loss from increasing customer size increases as illustrated in Figure A-2. The marginal gain from increasing the number of customers comes from the reduction in the tax burden that results as the fixed costs are spread over more customers. The marginal gain from reducing taxes is likely to be greatest when there are already large taxes in place.

Figure A-2



Therefore, the marginal gain from lowering taxes will tend to decrease with existing size. The intersection determines the optimal N .

When there is no cost to using a price system, then the firm can "tax" in all configurations of demand (i.e., in Q_1^L and Q_1^H separately) and that means that any total tax burden can be financed by causing less marginal distortions in any one state on average, than when there is a constraint that taxes in two demand states must be equal. Therefore, for any N , the marginal gain from financing - which depends on existing distortions - tends to be lower under competition. This lower marginal gain from financing is illustrated as the dashed curve in Figure A-2. The marginal loss from increasing N will tend to remain roughly the same for any N whether or not the price system is costless because of the perfect rationing assumption. As Figure A-2 indicates, this implies that the firm is smaller when use of the price system is costless.

Theorem 4: As rationing ability improves, ceteris paribus,

$$\frac{\partial t_0}{\partial R_0} > 0 \text{ and } \frac{\partial t_1}{\partial R_0} < 0, \text{ where}$$

$$R = R_0 [1 + a\rho] \quad a > 0$$

Proof: One of the first order conditions for the problem specified in (6) is:

$$t_0 Q'_0 = \lambda [t_0 Q'_0 + Q_0] \text{ or}$$

$$(A-4) \quad \tau_0 \epsilon_0 = \frac{\lambda}{1-\lambda}, \text{ where all terms were}$$

defined previously. As R_0 increases, the marginal deadweight loss from increasing t_1 increases. This causes the shadow price $(-\lambda)$ associated with the financing constraint to increase. From the above first-order condition, τ_0 increases as $(-\lambda)$ increases. As τ_0 increases, τ_1 falls to keep revenue constant.