MARKET BEHAVIOR UNDER UNCERTAINTY

by

DENNIS WILLIAM CARLTON

B.A., Harvard University
1972

M.S., Massachusetts Institute of Technology
1974

SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF

PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF

TECHNOLOGY

September, 1975

Signature of Author. . . . .
Department of Economics, July 22, 1975

Certified by . . . .
Thesis Supervisor

Accepted by. . . . .
Chairman, Departmental Committee on Graduate Students

TABLE OF CONTENTS

LISTING OF FIGURES

Abstract

MARKET BEHAVIOR UNDER UNCERTAINTY

by

Dennis William Carlton

Submitted to the Department of Economics on July 22, 1975, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

For many markets, the classical assumptions of competitive market behavior do not seem to apply. Prices do not adjust at each instant of time to keep supply and demand in balance, and firms never feel that they can sell all they want at the going market price. More realistic assumptions are that price responds to underlying market forces but does not vary at each instant, and that firms face a random demand for their product and are concerned with the risk of either overproducing or underproducing.

This thesis studies markets characterized by demand (or supply) uncertainty, price inflexibility and a lead time for production. Demand is uncertain for the period during which prices are inflexible. At the beginning of each market period, firms decide how much to produce, set prices and then observe their random demand. In terms of the conclusions of the analysis, it does not matter whether a firm decides how much to produce beforehand, just as long as there is some prior committment to production that must be made before the random demand can be observed. It appears that for many markets this description of market operation is more applicable than is the classical description which relies on a price mechanism which somehow instantaneously adjusts to always keep supply and demand in balance. Despite the realism of these assumptions of market operation, the resulting market behavior and its consequences differ drastically from what the classical analysis predicts.

For the markets under study, firms compete with each other on price and the probability of satisfying a customer. Equilibrium is defined for such markets, and its properties examined. In equilibrium, supply need not equal demand, the price will exceed the cost of production, and the probability that a customer will be unable to satisfy his demand will definitely exceed zero. The properties of this equilibrium are compared to those of the corresponding deterministic equilibrium.

Unlike the case of deterministic markets, it is not true that these competitive markets will lead to a socially desirable market operation. In general, the government will have to use its taxing powers to subsidize firms in order to maximize social welfare. A fair trade and monopoly equilibria are briefly examined, and it is found that a monopolist, and not price taking competing firms, is more likely to adopt policies that

will stabilize the smooth functioning of the economy in increasingly uncertain times.

The central focus of the analysis is on the transmission of uncertainty between firms and these firms' response to their uncertain environment. The stochastic nature of demand of one firm affects another firm's costs. The decision of firms to vertically integrate and produce some of the input for themselves affects the stochastic structure of demand in the input market. Price incentives are not sufficient to insure that firms will take account of the effect of their actions on the transmission of uncertainty to other firms. Private incnetives to vertically integrate are likely to exist, even though vertical integration can be socially undesirable. The strong incentives for a firm to vertically integrate arise because the vertically integrated firm is able to satisfy its high probability demand by itself and pass the low probability demand on to others. Competitive markets under uncertainty cannot be relied upon to properly allocate production and risk between interacting firms. However, prohibiting vertical integration solves one problem but creates another. It turns out that the ability of an integrated firm to better coordinate the characteristics of its own internally produced input (i.e. price and probability of availability) makes it more likely that an integrated firm, and not a nonintegrated one, will have an incentive to develop and introduce new and beneficial technology.

The models of this thesis are not alternatives to the classical model but instead are more general than the classical model and include it as a special case. Many characteristics of market behavior which are incomprehensible in the classical framework have a clear explanation when viewed in the more general framework of the models of this thesis. Supply not equaling demand, shortages, concern with obtaining assured supplies, incentives for vertical integration, the risk of under or overproducing or of not fully utilizing the firm's capital stock, all become natural features of market operation in the more general models. It is only by explicitly examining the effects of uncertainty on firms' responses that these features of market behavior together with their consequences can be fully comprehended.

Thesis Supervisor: Franklin M. Fisher

Title: Professor of Economics

ACKNOWLEDGMENTS

I wish to thank the members of my thesis committee, Professors Peter A. Diamond, Franklin M. Fisher, and Robert E. Hall for their many helpful comments and suggestions. I especially wish to express my thanks to the chairman of my thesis committee, Professor Fisher, for his excellent advice and constant encouragement during all phases of this research. His guidance and supportive comments, especially during that frustrating time when research is just beginning, are gratefully appreciated. While at M.I.T., I have had the good fortune of working with Professor Hall as a research assistant. My many conversations with him about the subjects in this thesis and about economics in general greatly shaped my thinking about several of the ideas expressed in this thesis. Conversations with Professors Bishop and Hausman, and my colleagues, Roger Gordon and William Keeton helped clarify my thinking on certain points. I am grateful to Rick Oiesen for providing two useful references. I thank Wilma Hawkins, Martha Yee and Katherine Crowley for their excellent typing.

Growing up in an extended family allows one the opportunity to develop close relationships with many family members. I wish to thank them all for the devotion and help they have provided me during my years of schooling. My parents, especially, I thank for their willingness to make any sacrifice, financial or otherwise, to insure that I would have an opportunity to attend college and then graduate school. My final thanks are reserved for my wife who provided editorial assistance and general support. Her constant encouragement during times when things looked especially difficult always kept my spirits up. Next year, it's my turn to be the family cheerleader, as my wife finishes up her thesis.

To my darling wife, Janie

CHAPTER 1

Introduction

Most economists would agree that the large majority of markets do not precisely fit the classical assumptions of competition. For many markets, prices do not adjust at each instant of the day to balance supply and demand. Moreover, firms often do not know how much of their product will be demanded each day.

There are good reasons why most markets depart from the strict classical assumptions. Changing prices frequently is time consuming and may be costly. More importantly, prices may have to remain in effect for some time if their "signal" is to be received. The demand that an individual firm sees is random because the number of customers that frequent the firm will generally vary from day to day. In formulating its operating policy, a firm must take into account the randomness of its demand. Firms do not feel that they can sell all they want at the going market price and are concerned with overproducing and being left with unsold goods. Firms are also concerned with underproducing and they stock inventories to guard against the possibility of losing a potential sale. In these markets, it is an outcome of the market process that occasionally some customers will be unable to purchase the good. Surely everyone has had the experience of going to a store, only to find that the last jelly bean had just been sold at a price below that which he would have been willing to pay.

For these uncertain markets, the amount that a firm is willing to supply depends not only on the going market price, but also on the entire stochastic structure of demand that it faces. In this environment, supply cannot be defined without first specifying the random structure of demand. The Marshalian separation between demand and supply disappears, and a more sophisticated analysis is required which explicitly incorporates the role of uncertainty into the decisions of market participants.

There will be three essential features of market operation that we will study; price inflexibility, demand uncertainty, and timing considerations. By price inflexibility, we do not mean that prices do not respond to permanent shifts in the underlying supply and demand factors, but only that prices cannot be adjusting at each instant of time. An important feature of the analysis will be to determine exactly how prices are endogenously determined by market forces. Demand uncertainty means that, at the beginning of any market period, after prices have been set, firms do not know for sure what their demand will be, although they do know what the random distribution of demand looks like. Demand is uncertain over the period for which prices are inflexible. Timing considerations refer to the need to have produced (or to have made some prior committment to production such as the purchase of equipment) before the unknown customer demand is observed.

For markets with these three characteristics, it will be a natural feature of the market to have some customers being unable to purchase the good, and some firms being unable to sell all of their stock. Customers will have preferences not only for the price of the good, but also for the probability of obtaining it. The "customers" can also be interpreted as being other firms who are trying to buy factor inputs for their production process. With this interpretation, we obtain a model where it is perfectly natural for firms to be concerned with obtaining an "assured" supply of the input, a concern that appears uppermost in the minds of businessmen.[1]

[1] See A. Chandler, Strategy and Structure: Chapters in the History of American Industrial Enterprise, M.I.T. Press, 1964, Ch. 1.

In sharp contrast, in the classical model of supply and demand, notions of shortage and hence of assured supply make no sense. It is only in this more general model that it is possible to examine the reasons for and consequences of the incentives for achieving an assured supply.

It is not immediately clear what the consequences of these three non-classical features of market operation are, even though these three features would appear to be realistic characterizations of many market operations. How do firms compete in such markets? Can equilibrium be meaningfully defined and if so how does it compare to the classical equilibrium when the uncertainty is removed from the demand side? Will this equilibrium be Pareto-Optimal? Would society benefit if the government paid lump sum subsidies to firms so as to encourage them to expand their production of the good?

Suppose we interpret customers as other firms trying to purchase inputs for their production process. How do firms interact and transmit their uncertainty between each other? What are the consequences of this transmission of uncertainty? Since one firm's decision affects the entire stochastic structure of demand (or supply) facing another firm, can we expect the free market to insure that competing firms take full account of their actions on other firms? What incentives are created when firms face an uncertain input supply? Will the firms have an incentive to produce the input for themselves - in other words, will firms vertically integrate through internal growth? Is such vertical integration socially desirable, or should it be prohibited? How does the choice of a production technology depend on market structure in this uncertain environment?

This thesis will address the above questions and many related questions in an attempt to trace through the consequences of the three non-classical, but realistic, assumptions of market operation. Answers to many of the questions will differ from what one might have expected from an extrapolation of classical precepts. One main finding is that for the markets under study, firms do not take full account of the effects of their action on others. Free markets will not in general lead to a market structure or operating policies that are desirable from society's point of view.

Outline of Thesis

The next chapter investigates how a single market characterized by demand uncertainty, price inflexibility and timing considerations operates. A simple model is presented to try to capture the essential features of market operation. In the model, firms set price and decide on production at the beginning of the market period, then they observe their random demand. Firms are concerned with underproducing and losing potential sales, and with overproducing and incurring extra holding costs. Consumers have preferences for both the price and the probability of obtaining a good. It is shown how an equilibrium will be established if there are several firms competing with each other, and if the random distribution of demand is the same each period for each firm. Equilibrium is characterized by both a price and a probability of obtaining the good. In general, supply will not equal demand in equilibrium. We prove that as the number of customers increases, the equilibrium price approaches the price that would prevail in the corresponding deterministic market, and the equilibrium probability of obtaining the good approaches one. In equilibrium, the percent discrepancy between supply and demand, goes to zero as the number of customers increases. However, the absolute discrepancy between the equilibrium supply and demand becomes arbitrarily large as the number of customers increases.

In Chapter 3, we see how fair trade pricing and monopoly pricing affect the market equilibrium. We compare these equilibria to that derived in Chapter 2. The fair trade equilibrium is found to be non Pareto-Optimal. We show that for fixed prices the response of a monopolist to an increase in uncertainty is more likely to allow an economy to function smoothly than is the response of individual competing firms.

In Chapter 4, we investigate the welfare implications of these uncertain markets. It turns out that none of the previously discussed market equilibria, in general, lead to the socially optimal point. The social optimum will, in fact, involve paying lump sum subsidies to encourage firms to expand their production of the good.

Finally, in Chapter 5, we take up the important question of firm interaction under uncertainty. We suppose that customers are other firms trying to purchase inputs for their production process. The desire to obtain an assured supply of the input creates an incentive for a firm to vertically integrate and thereby produce some of its own input. Because demand is random, firms which produce their own input run the risk of having unused input at the end of the market period. When firms buy their input from other firms, it is the other firms who bear this risk. It turns out that free markets cannot be relied upon to achieve the socially desired allocation of risk and production between firms. There is an externality involved since the behavior of one firm affects the entire stochastic structure of demand (or supply) that other firms see. For the markets studied in Chapter 5, we show that it is socially undesirable to allow vertical integration to occur through internal growth, but that there exist strong private incentives for such vertical integration to occur. Another finding of Chapter 5 is that a new technology which could benefit society is more likely to be introduced in a market structure involving vertical integration than in one involving no vertical integration. We are led to a Schumpeterian view that some inefficiency caused by market structure must be tolerated in order to create an environment in which new and beneficial technology is likely to be developed and rapidly introduced.

A heuristic explanation of the main results is given in the text so that the reader, not interested in the technical details of the proofs, can omit the proofs, yet still understand why the results are true. Technical appendices contain some of the more tedious mathematical proofs of the results.

CHAPTER 2

Competitive Market Clearing with Demand Uncertainty and Price Inflexibility

2.0 Introduction

There is a large literature on the effects of uncertainty on firm behavior.[1] Analyses of competitive markets focus on the effect of having uncertainty in price and maintain the assumption that firms can always sell all they want at the future uncertain market price.[2] There are never any shortages in equilibrium. In his pioneering works, Mills[3] has examined the effect of demand uncertainty and price inflexibility on the behavior of a monopolist who must decide what price to charge and how much to produce before demand can be observed. Surprisingly, despite the realism of the assumptions of demand uncertainty, price inflexibility, and a lead time necessary for production, there has been no attempt to examine the implication of these assumptions within a competitive environment. The purpose of this chapter is to provide such an examination, and to derive and investigate the properties of an equilibrium in which it is natural to have supply not equal to demand.

This chapter investigates the behavior of competitive market operation when there is demand uncertainty, price inflexibility and a lead time necessary for production. A model is developed to illustrate the distin-

---

[1] See M. Rothschild, "Models of Markets Organization with Imperfect Information: A Survey", Journal of Political Economy, 1973 and J. McCall, "Probabilistic Microeconomics", Bell Journal of Economics and Management Science, 1971 and the references cited therein.

[2] See, for example, E. Zabel, "A Dynamic Model of the Competitive Firm", International Economic Review, 1967.

[3] E. Mills, "Uncertainty and Price Theory", Quarterly Journal of Economics, 1959, and Prices, Output and Inventory Policy, Prentice-Hall, 1962.

guishing features of the behavior of such markets. The model forms a basis for much of the subsequent analysis in later chapters. It represents a logical and necessary first step in the analysis of the effect of the transmission of uncertainty between firms in different markets.

The model postulates conditions that would appear to be in accord with the operation of many actual markets, namely, that firms must decide on price and production[1] before demand can be observed. For such markets, the notion of a supply curve is not a useful concept and cannot a priori be defined in a reasonable fashion. In the model, each good will have two characteristics associated with it, namely its price and the probability that it can be purchased. Firms will compete amongst themselves until an equilibrium is reached. Market clearing will require equilibrium along the dimensions of both price and probability of obtaining the good. In equilibrium, supply will not, in general, equal demand and there will always be some customers who are unable to purchase the good. As the number of customers increases, the equilibrium price approaches its value in the corresponding deterministic market, and the probability of shortage falls to zero. The percent discrepancy between supply and demand goes to zero as the number of customers increases. However, in general, the absolute discrepancy between supply and demand will increase as the number of customers increases, so that equilibrium can involve having an arbitrarily large number of customers being unable to purchase the good.

2.1 The Model

There are N identical firms who compete with each other, and L <u>identical customers</u> each of whom has a nonstochastic demand for the good given

[1] In terms of the conclusions of the analysis, it is only necessary that there be some prior committment to production (e.g. the purchase of equipment) that must occur before demand can be observed.

by x(p). To make the assumption of competition plausible, the number of firms N will be considered to be large enough to prevent firms from having any monopoly power. The number of customers is assumed to be considerably larger so that the number of customers per firm, L/N, is reasonable (i.e. over 35). Individuals maximize expected utility and firms maximize expected profits.

At the beginning of each period, each firm sets price, which remains in effect for the entire period, and decides how much of the good to stock for the period. No deliveries of the good can occur during the period. The cost per unit of the good is c, where c is strictly positive. We assume that the good is perishable so that it is impossible to hold inventories between periods. A firm that is left with unsold goods at the end of the period must throw them away[1].

During each period, each of the L identical customers frequents a firm of his own choosing. If a customer finds a firm out of the good, he simply leaves the store and does not obtain the good for that period. He does not search at the other stores[1]. Buyers have preferences for not only how much they purchase and spend on the good, but also for the probability of being able to buy the good. Therefore, competition does not force firms to necessarily charge the same price but rather to offer price-shortage combinations which leave the consumer at the same level of expected utility.

Equilibrium in an uncertain market is said to exist when 1) consumers are indifferent as to which of the firms they shop at each period, and 2) no firm, behaving optimally, can offer a price-shortage combination which would leave all consumers better off, and which would allow the firm

[1]In a later section, we discuss why the holding of inventory and consumer search behavior would not alter any of the qualitative features of the model.

to earn non-negative expected profits. Before examining how market equilibrium is determined, let us first look at the incentives facing individual consumers and firms.

2.2 Consumer Behavior

In the model, a consumer visits one firm each market period to try to satisfy his demand for the good. Before visiting a firm, the consumer does not know whether the firm has any goods left to sell. Instead, the consumer has an idea of the price this firm charges and the probability that this firm can satisfy his needs. If the customer finds the firm sold out of the good, then he must do without the good for that period, and must spend his money on an alternative good which we assume is always available at a price of one. If the firm is not sold out, then the customer buys the good at the price charged, according to his deterministic demand schedule $x(p)$. The preferences of the consumer do not change over time. Since the good is nonstorable, the fact that a customer was unable to purchase the good last period will not affect his demand this period.

In calculating his expected utility from going to any firm, a customer is concerned with both the probability, $1 - \lambda$, of obtaining the good and the price, p, charged for the good. We can write his expected utility as $U(1 - \lambda, p)$. The function U defines the isoutility contours between $1 - \lambda$ and p that leave a consumer indifferent. Typical isoutility contours are drawn in Figure 2-1.

The diagram shows that along any isoutility curve, as price rises, the probability of satisfaction must rise if consumers are to remain indifferent. Also, for any fixed probability of satisfaction, consumers always prefer lower prices.

Figure 2-1 - Isoutility Contours

Consumers will always try to reach their highest isoutility contour, and will only go to a firm that they think will provide this highest isoutility level. If the buyer believes that several firms provide this highest utility level, then he will choose among them randomly.

Suppose the consumer has a Von Neumann utility function that expresses his tastes for risks and for the good x and the alternative good z. Call this utility function $u(x, z)$. Let the income of the consumer be denoted by Y, and let $V(p, Y)$ be the associated indirect utility function so that $x(p) = (-1)^{V}p/V_{Y}$. Then, $U(1 - \lambda, p) = (1 - \lambda)V(p, Y) + \lambda u(o, Y)$. From this expression for $U(1 - \lambda, p)$, it is easy to show that it is impossible to determine whether the isoutility curves drawn above should be convex or concave (i.e., $\frac{d^2 1-\lambda}{dp^2} > 0$ or $\frac{d^2 1-\lambda}{dp^2} < 0$). It is possible to construct examples of either case. Experimentation with several utility functions has indicated that the faster the marginal utility of good x declines, the more likely is the isoutility curve to be concave. The smaller the difference between $V(p, Y)$ and $u(o, Y)$ the more likely is the isoutility curve to be convex. In general, no strong conclusion about the shape of the isoutility curves seems justified, although based on the simple examples, it appears that the case of concave isoutility curves is more likely.

Because of the ambiguity in determining the shape of the isoutility curves, the subsequent results will not in general depend on any assumed convexity or concavity properties of isoutility curves. Instead we will only require the very weak assumptions that the isoutility curves exist over the relevant range[1] in $(1 - \lambda, p)$ space, that they are continuous, and that they satisfy an upper and lower Lipschitz condition. This latter condition postulates that there exist two numbers, b and B, such that $0 < b < B < \infty$ and such that the slope along any isoutility curve always lies between them. This condition rules out horizontal and vertical segments for isoutility curves, which, as will be shown later, can result in pathological and uninteresting market behavior. Basically, the Lipschitz requirements insure that the consumer is never willing to make infinite trade-offs in either the p or $1 - \lambda$ directions.

2.3 Behavior of the Firm

Since consumers will wind up going only to those firms that provide the highest utility level in the market, competition forces firms to take the utility level as given. (If instantaneous production were possible so that no shortages could occur, then each good would have only one characteristic, price, associated with it. In that case, utility-taking behavior is equivalent to price-taking behavior.) At the beginning of each period, firms have to decide on a price and stocking policy so as to maximize their profits subject to the constraint that they provide at least the given level of utility to consumers. Firms know that if they remain competitive with the other firms, then they will randomly receive $\frac{1}{N}$th of the total population L.

---

[1]By this assumption we simply mean that there is some range of prices, which includes $p = c$, the cost of production, for which the consumer is interested in purchasing the good. In other words, if the consumer does not have positive demand for prices near c, then the market for the good will not exist, and there is nothing to analyze.

We can write the total amount that the firm decides to stock at price p as $s \cdot x(p)$. The variable s can be interpreted as the maximum number of customers that a firm can satisfy that period. Henceforth, we will refer to s as customer capacity. Clearly, the amount that a firm decides to stock affects the probability that a customer will be able to obtain the good from that store.

Let us examine the relation between the expected number of customers, M, who will find the firm out of the good, and the customer capacity, s, that the store provides. Let pr(i) stand for the binomial probability that i customers from the L customers arrive at the firm. Then, we can write that

$$M(s) = \sum_{s+1}^{\infty} (i - s)pr(i)$$

If all N firms follow the same operating policies, then the total expected number of customers who will be dissatisfied is $N \cdot M$, and the fraction of dissatisfied customers will equal NM/L. The fraction $1 - \lambda$ of customers who are able to obtain the good can be written as

$$1 - \lambda(s) = 1 - \frac{N \cdot M(s)}{L} \qquad (1)$$

In the appendix to this chapter, we show that using the normal distribution to approximate the discrete binomial process of customer arrival, the probability of satisfaction function, $1 - \lambda(s)$ can be written as:

$$1 - \lambda(s) = \frac{\sigma\ I(u) + s}{\sigma^2} \qquad (2)$$

where $\sigma^2 = L/N$, $I(u) = \int_{-\infty}^{u} [t-u]f(u)du$, $f(u)$ = normal density function, and $u = \frac{s-\sigma^2}{\sigma}$.

Technically, as derived above, the $1 - \lambda(s)$ function applies to an individual firm only when all firms follow the same operating policies. However, if customers and firms calculate the probability of satisfaction that an individual firm offers as the expected shortage of that firm divided by the expected number of customers for that firm, then we can interpret (1) as applying to the individual firm.

More importantly, since $M(s)$ and $1 - \lambda(s)$ are in one to one relation by (1) the entire analysis could be carried out in (M, p) space and not $(1 - \lambda, p)$ space. Since M obviously applies to the individual firm, this approach would avoid any questions about whether the derived curves apply to individual firms. In such an approach, each firm is regarded as choosing an expected shortage, M, and price, p, combination, and consumers are regarded as having preferences for the price and expected shortage of each firm. Because of the one to one relation between $1 - \lambda$ and M in (1), the result of such an analysis will be identical to one in $(1 - \lambda, p)$ space. However, it seems more natural to talk of consumers as having preferences for the probability of satisfaction, $1 - \lambda$, and not the expected shortage, M. For these reasons, we carry out the analysis in $(1 - \lambda, p)$ space, and regard the $1 - \lambda(s)$ curve as applying to individual firms. Furthermore, we show later that it is reasonable to expect that all firms (remember all firms are identical) will follow the same operating policies in equilibrium, so that any remaining misgivings about interpreting the probability of satisfaction function $1 - \lambda(s)$ as applying to individual firms completely disappears in equilibrium.

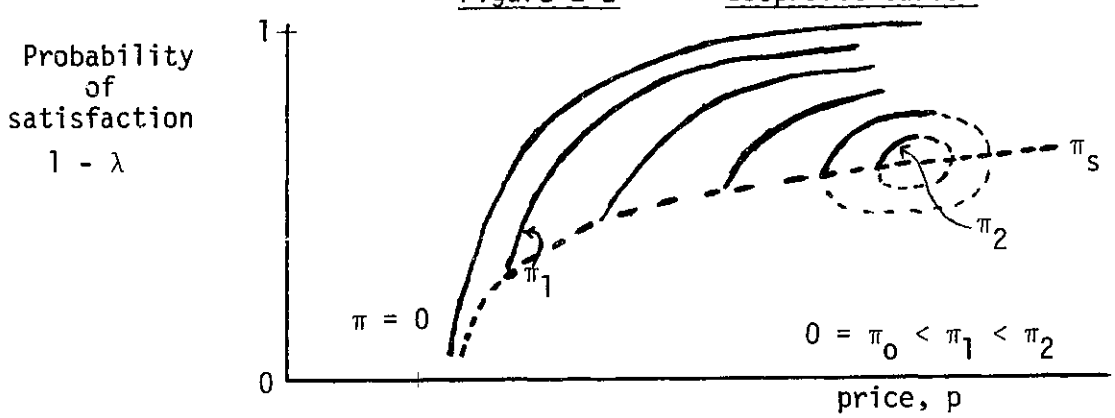
For the given level of utility, firms want to choose a price, p, and a customer capacity, s, so that profits are maximized and the consumer is able to achieve the given level of utility. When firms remain competitive by offering the given level of utility, they randomly receive their equal share of the L customers. Letting pr(i) stand once again for the probability that i of the L customers visit a firm this period, we can write that expected profits equal

$$\pi(s,p) = p\cdot x(p) \sum_{0}^{s} i\ pr(i) + px(p)s \sum_{s+1}^{L} pr(i) - csx(p) \qquad (3)$$

The first term in (3) represents expected sales revenue when $i \leq s$ customers come to the firm, while the second term represents expected sales revenue when more than s customers come to the firm. The last term (3) is the cost of being able to service s customers. Since (2) expresses a one to one relation between the probability of satisfaction $1 - \lambda$ and the customer capacity s[1], we can interpret (3) as expressing profits as a function of $1 - \lambda$ and p.

Regarding profits as a function of $1 - \lambda$ and p, we can draw isoprofit curves in $(1 - \lambda, p)$ space. A typical family of such curves is depicted below.

Figure 2-2 - Isoprofit Curves



[1]It should be obvious that the probability of satisfaction and the customer capacity are in one to one relation. Mathematically, this is so since, as proved in the mathematical appendix to this chapter, $\frac{d(1-\lambda)}{ds} > 0$.

The two isoprofit curves at the far right of the diagram are drawn to illustrate that each isoprofit curve involving positive profits "turns around" on itself as price rises sufficiently high to drive demand to zero. Since consumers always prefer to be on the northwest boundary of the isoprofit curves, competition will insure that the "dotted" segments of the isoprofit curves are never observed. The heavy dotted line in the diagram represents the π_s curve which is derived by setting $\frac{\partial \pi}{\partial s} = 0$ in (3). As the diagram illustrates, isoprofit curves cross the π_s curve vertically, and so the relevant portions of all isoprofit curves emanate from the π_s curve. For any fixed probability of satisfaction, profits increase as price increases. Hence, in the diagram $\pi_1 < \pi_2$. The curve on the far left of Figure 2-2 represents the zero profit curve.

For any given isoutility level, $\bar{u}$, the firm will choose to operate at the point of tangency between the isoutility curve representing isoutility level, $\bar{u}$, and the highest isoprofit curve. No firm ever chooses to operate to the left of the $\pi = 0$ curve since that represents negative expected profits.

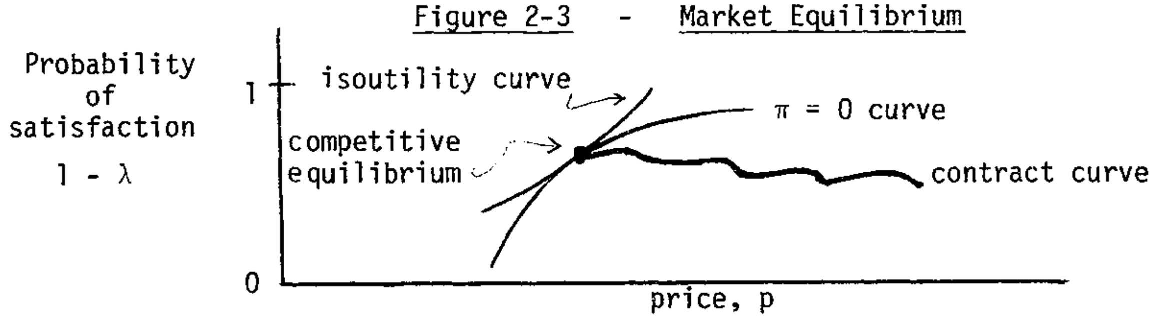
2.4 Market Equilibrium

In the diagram of the isoprofit curves, superimpose the isoutility curves of consumers. We can define a contract curve as the locus of tangencies between the isoutility and isoprofit curves. Firms always operate on this contract curve.

In a classical market, firms compete with each other by offering to consumers lower prices (i.e., higher utilities) than other firms. Prices (or consumer utilities) continue falling (rising) until firms have no in-

incentive to lower price (raise utility) any more. Analogously, in this market, firms compete with each other by offering better (i.e., higher utility) combinations of price and probability of satisfaction to consumers. The utility level is "bid" up until there is no incentive for any firm to continue to alter its price-probability of satisfaction combination. This point occurs when the contract curve intersects the zero profit ($\pi = 0$) curve. At this point, firms would prefer to go out of business rather than offer a higher utility combination to consumers and earn negative expected profits. Hence, competition on the utility level forces the market equilibrium up the contract curve, until the zero profit curve is reached. Equilibrium can be regarded as the tangency[1] between the zero profit curve and the highest attainable isoutility curve. This equilibrium is depicted below.

Figure 2-3 - Market Equilibrium



Before investigating the properties of the market equilibrium defined above, let us consider the competitive process in a little more detail. Firms are assumed to be "utility-level" takers, yet in the description of how a market reaches equilibrium, we stated that firms "compete" with each other on the offered utility level. If firms take the utility level as given, which firms are changing the utility level in the approach to market equilibrium? The problem here is identical to the one in pure com-

[1]We will soon argue that corner solutions are uninteresting, unlikely, and under the Lipschitz Assumptions on preferences, impossible.

petition. If all firms are price takers, who ever changes price to insure that price clears the market? Traditional explanations rely on a Walrasian auctioneer. More ambitious attempts at realistic adjustment mechanisms have met with little success. "Despite great and admirable efforts by many leading theorists, we have no...satisfactory theory of how equilibrium is reached."[1]

In the next section, a story is presented to justify reaching the equilibrium of Figure 2-3 without introducing a Walrasian auctioneer. The story tries to capture what the essential features of this competitive process are, but stops short of presenting a mathematically detailed dynamic analysis of consumer and firm decisions over time. The story is intended to suggest that competition will result in a stable market operation that leads to the equilibrium discussed above.

2.5 A Story of Competitive Market Operation

In the story below, we try to incorporate several features that describe the behavior of markets characterized by demand uncertainty and price inflexibility and a lead time necessary for production. At the risk of repetition, let me reemphasize that the story is simple and is designed solely to suggest that it is reasonable for competition to lead to the equilibrium described in the previous section for the markets under study. The story is not meant to solve the problem of dynamic adjustment to equilibrium in markets characterized by demand uncertainty and price inflexibility. Indeed, in view of the failure to date of establishing a dyna-

[1] F.M. Fisher - "Quasi-Competitive Price Adjustment by Individual Firms: A Preliminary Paper" - Journal of Economic Theory, 1970, p. 195.

mic theory for the somewhat simpler classical markets, it would undoubtedly require another thesis, not just a section, to even begin to treat the problem in its full complexity.

First, consider consumers. Consumers do not have perfect information at every instant of time, but they eventually do acquire all the information in an unchanging market. By these assumptions, we mean that a consumer can get a bad deal at some store because he was not aware that another firm had just decided to offer a better deal. However, eventually information will flow to the consumer (e.g. he exchanges information with his friends) so that he will find out that this other store is offering a better deal. The consumer will not consider returning to the original store unless he becomes convinced that it will offer as good a deal as this other store. This may mean that the original store has to spend money on advertising or other promotional schemes to get the customer back in his store and reestablish his competitiveness in the mind of the consumer.

Now consider firms. We expect that firms that offer the best deals in the marketplace are eventually rewarded, while firms that offer poor deals are penalized. As just described, we can view firms who offered poor deals as having to go to some added expense in order to reestablish their competitiveness in the minds of consumers. However, what must be avoided is to have instantaneous and infinite jumps in demand for firms as they marginally alter their operating policies. This is rarely observed in actual markets, and would lead to a chaotic market process if it ever did occur. Incorporating all these features, let us now construct a simple story of how the markets under study would reach a competitive equilibrium.

A market period will consist of one day. Firms maintain the same price-stocking policy throughout a one week period. Initially, consumers know nothing about the firms, and so during the first week, each customer randomly frequents one store per day to try to satisfy his demands. After the final bell has rung on Friday to end the market trading for that week, it becomes known which firms were offering the best deals. Firms which did not offer the best deals will receive zero demand all next week, unless they do something to convince consumers that next week they will be competitive with other firms.

Between Friday night and Monday morning, firms which did not offer the best deals in the market have to spend money on some sort of promotional scheme to reestablish their competitiveness in the minds of consumers. They must convince consumers that next week, they will offer deals that are at least as good as the best one that was offered this week. Firms must make good on their promise or else customers will never believe them and return to them again. Firms do not necessarily advertise the precise policy they will adopt next week[1] - or if they do, consumers discount such information. Otherwise, we would have the problem of instantaneous and infinite jumps in demand, a situation that leads to market chaos. Instead, stores are viewed as spending money to create the impression in consumers minds that they will be competitive with the other firms. There are many ways to envision the

[1] Firms might not even know what policies they will adopt at the time they have to advertise. For example, we can imagine that one hour before the Friday market ends, only the best firms are notified who they are. Before they close up shop for the weekend, unnotified stores erect signs saying that whatever the policies of the best store were this week, they will do at least as good next week. Everyone in the Boston area is familiar with an advertising policy of this sort, in which a flamboyant car dealer rants that he offers the best deals in town and will never be undersold-if a customer can find a car on his lot that is sold at a lower price elsewhere, he will give the customer the car for free.

form that these promotional campaigns can take from erecting billboards to giving out candy bars. Once the competitiveness of each store is reestablished in the minds of the consumers, next week consumers randomly frequent the stores once again, and the process is repeated.

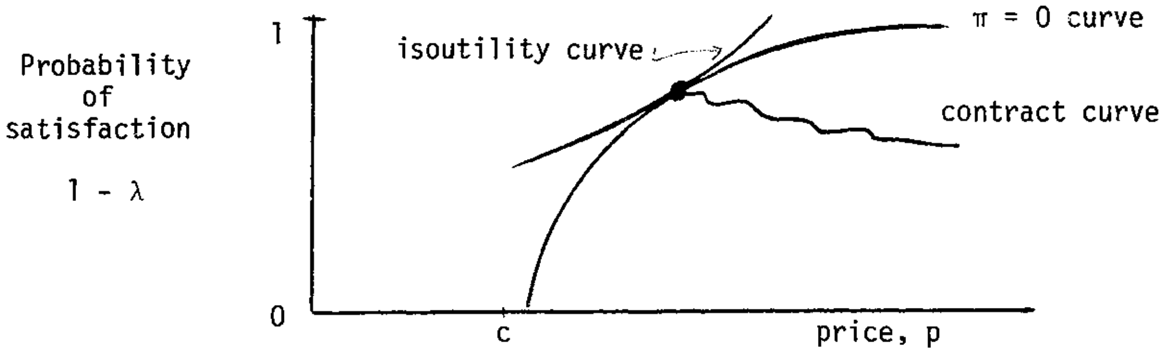
The important features of the above story are that firms who remain competitive in the minds of consumers expect to receive a random equal share of the market. Firms never believe that they can obtain an instantaneous and infinite increase in their demand by altering their operating policies. Firms have an incentive to offer the best deal in the marketplace so as to avoid the cost of reestablishing their competitiveness in the minds of the consumers. We assume that this cost is sufficiently large so that it wipes out any gains that could be made by behaving like a monopolist for one week.[1] (For example, the cost of reestablishing competitiveness could depend on just how bad a deal a firm offered this past week.)

We can now see how equilibrium is reached. After the first week is over, firms know what level of utility they must at least offer for next week. As seen in the previous section, firms will always operate on the contract curve defined by the points of tangency between the isoutility and isoprofit curves. Firms have an incentive to move along the contract curve and offer a level of utility slightly higher than that offered last week in order to try to avoid the costs of reestablishing their competitiveness in the minds of consumers, next week. As long as expected profits remain positive, firms will have an incentive to "bid up" the offered utility level. Once expected profits become zero, however, firms would prefer to go out of business rather than offer a higher level of utility. Utility level competition in this market replaces the price competition (which also can be viewed as utility

[1]We also assume that firms desire to remain in business for more than a week.

level competition) of the classical market. Competition forces the market equilibrium up the contract curve until the zero profit curve is reached, as depicted below.[1]

Figure 2-4 - Market Equilibrium



2.6 The Properties of Equilibrium

2.6.1 General Properties

The distinguishing feature of market clearing in the markets under study is that equilibrium involves both a price and a probability of satisfaction. Consumer preferences for both price and probability of satisfaction are needed to determine equilibrium. There are three distinct quantity variables - the amount supplied, demanded, and sold - none of which need equal each other in equilibrium. Total supply can fall short of total demand, yet there need not be any market forces to provide an incentive for firms to increase their supply. Consumers can prefer to take the risk of not being able to obtain the good rather than face a higher price for the good. For the markets under study, it is natural in equilibrium to have some customers unable to purchase the good. There is absolutely no reason to expect that in equilibrium the total amount supplied should equal the total amount demanded.

[1]With instantaneous production, the model becomes identical to the classical supply and demand model. For that case, the π=0 curve is a vertical line at p=c, and equilibrium as defined above, coincides with the classical equilibrium of price = c, probability of satisfaction = 1. We see then that the classical model is a special case of this model.

For the markets under study, it is obvious that a simple supply equals demand analysis will not suffice. Such an analysis fails because it ignores the uncertainty in the market place, and consumer's preferences for this uncertainty. A deterministic analysis would predict that the price of the good is c, the probability of shortage is 0, and the total amount demanded and supplied is L·x(c). As seen in the previous section, equilibrium will always involve a price greater than c,[1] a probability of shortage below 1, an amount demanded below L·x(c) (assuming that demand declines as price increases), and usually the amount supplied and demanded being in imbalance. Moreoever, in markets characterized by demand uncertainty, price inflexibility and a lead time in production, it is impossible to even define what is meant by a traditional supply curve. The amount that a firm is willing to supply at any price depends on the entire stochastic structure of demand.

2.6.2 Qualifications

In the previous sections, we always refer to equilibrium as the tangency between the highest isoutility curve and the zero profit. This statement needs some qualifications. If we want to examine the behavior of a particular market, it is useful to rule out certain types of behavior as unrealistic or uninteresting. For example, in Section 2.2 we assumed that for prices near c, per capita demand was positive. Without this assumption, the market will not exist, hardly an interesting case to study.

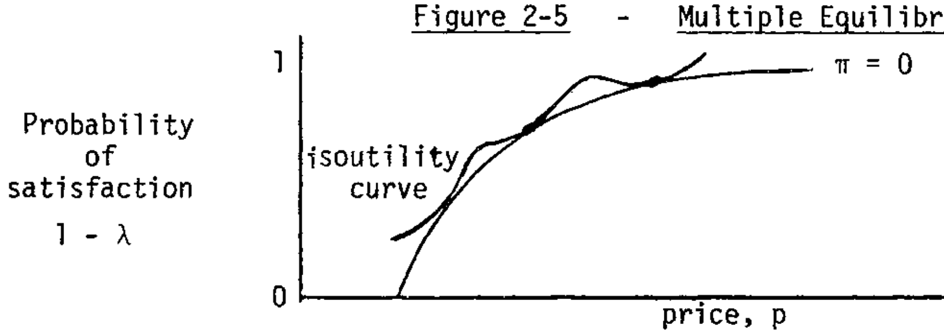
There are two non-tangency "equilibrium" points that are possible for the markets under study. Both these "equilibria" which correspond to "corner" solutions between the isoutility and zero profit curves are strange and/or uninteresting. If the isoutility curves are vertical, then equili-

[1] Price exceeds c, since price must cover not only the unit cost of production but also the cost of unsold goods.

brium involves zero production - again the uninteresting case of the market not existing. If the isoutility curves are horizontal, then equilibrium involves a very high price and each firm stocking enough of the good to by itself satisfy the entire market. Both these corner equilibrium seem sufficiently uninteresting to exclude them from further analysis. (In the appendix to this chapter, we show that the zero profit curve has a very large slope at the low price end, and a flat slope at the high price end. Hence the Lipschitz conditions on consumer preferences are sufficient to rule out the uninteresting corner solutions.) Henceforth, we assume that the market does indeed exist, and rule out the uninteresting case of a corner equilibrium occuring at the upper price range.

Finally, since we have not made any restrictive assumptions on the shape of the isoutility curves, it is obvious that it is possible to have multiple equilibria as depicted below.

Figure 2-5 - Multiple Equilibria



The dynamic adjustment process, as well as the initial starting point, would influence which point (s) the market winds up at. Although possible, the case of having some firms operating at one equilibrium point and the remainder at other equilibrium points does not seem particularly likely or interesting. Since all firms and all consumers are identical, it seems more reasonable to expect that all firms will wind up operating at the same equilibrium point. For the remainder of the analysis, whenever all firms and all consumers are assumed to be identical we will not discuss the possibility of having several equilibria simultaneously being in existence.

We now want to examine how the market equilibrium behaves as the customer per firm ratio, L/N, increases. This examination will clarify the relation between market clearing under certainty and under the uncertain conditions under study here. Since equilibrium is determined by the tangency between the isoutility and zero profit ($\pi = 0$) curve, it is necessary to establish some properties of the zero profit curve in order to understand how the equilibrium behaves as the customer per firm ratio increases.

2.6.3 The Zero Profit Curve

The properties of the zero profit curve play a key role in determining the behavior of equilibrium as the customer per firm ratio increases. In this section, we describe the relevant properties of the zero profit curve which are proved in the appendix to this chapter.

Since we have ruled out as uninteresting the case where the market vanishes (i.e., x(p) = 0), we can use (1) to write the condition for zero profits as

$$0 = \pi = p \cdot \sum_{0}^{s} i \; pr(i) + p \cdot s \sum_{s+1}^{L} pr(i) - cs$$

where all notation was defined previously. Notice that as a consequence of the assumption of a constant cost, c, the per capita demand x(p) does not appear in the zero profit condition. Using the normal distribution to approximate the binomial, we obtain that the zero profit condition can be written as

$$[\sigma I(u) + s] \cdot L \cdot p - c \cdot s \cdot N = 0 \qquad (4)$$

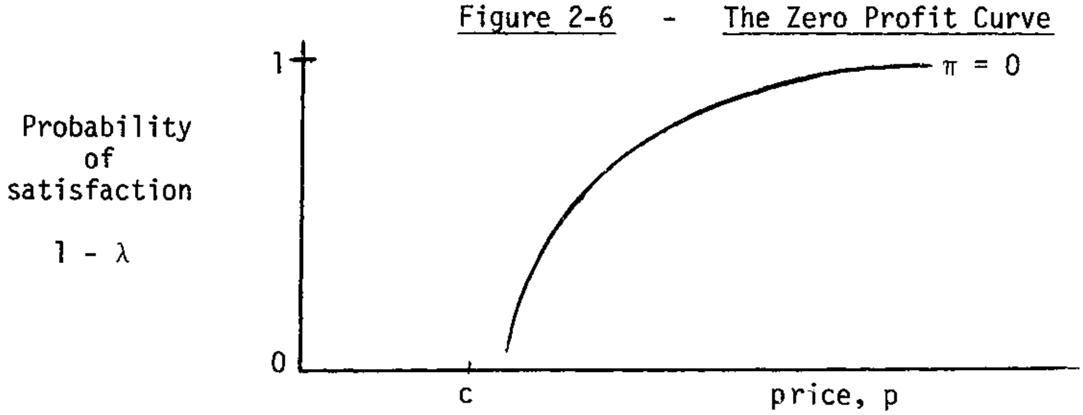
where all notation was defined previously beneath (1) and (2). Since the customer capacity, s, and the probability of satisfaction, $1 - \lambda$, are in one to one correspondence by (2), we see that (4) can be interpreted as expressing a relation between $1 - \lambda$ and p that must hold along the zero profit curve. We can write (4) as

$$\pi(1 - \lambda, p) = 0 \qquad (5)$$

There is a minor technical point associated with (4) and (5). Since we are using a continuous random variable to approximate a discrete positive random variable, there is a slight error involved. In particular, since the number of customers is bounded between 0 and L, we know that if customer capacity, s, for each store equals, L, then the probability of shortage equals 0. The continuous approximation would not necessarily tell us that this probability is exactly one, but only that it is very close to one. By the Central Limit Theorem, we know that any such approximation errors become insignificant for even moderate (i.e. 15-20) values of the customer per store ratio, L/N. In the subsequent analysis, we shall ignore such approximation errors.

The general shape of the zero profit ($\pi = 0$) curve is depicted below.

Figure 2-6 - The Zero Profit Curve



The $\pi = 0$ curve is concave (i.e. $\frac{d^2(1-\lambda)}{dp^2} < 0$), starts off with a very large slope at a point a little to the right of $p = c$ on the horizontal axis, rises to 1 as price increases, and has a very flat slope for sufficiently high prices. The curve always lies to the right of the vertical line $p = c$, since price must cover not only production costs, but also the cost of unsold goods. As price rises, firms can afford to provide a larger customer capacity, s. Hence along the $\pi = 0$ curve the probability of satisfaction increases to 1, as price increases.

The $\pi = 0$ curve begins a little to the right of $p = c$ on the horizontal axis, because for any finite customer per store ratio, $\frac{L}{N}$, there is always some probability that a firm will be unable to sell all its stock, even if it stocks only one unit of the good. The price along the horizontal axis at which the curve begins to rise moves closer and closer to c as L/N increases. Basically, as L/N increases, the probability of selling that first unit of the good approaches 1[1], so that if a firm produces only one unit, there will almost surely not be any unsold goods; hence, price need cover only the production cost, c, if the firm is to make zero profits.

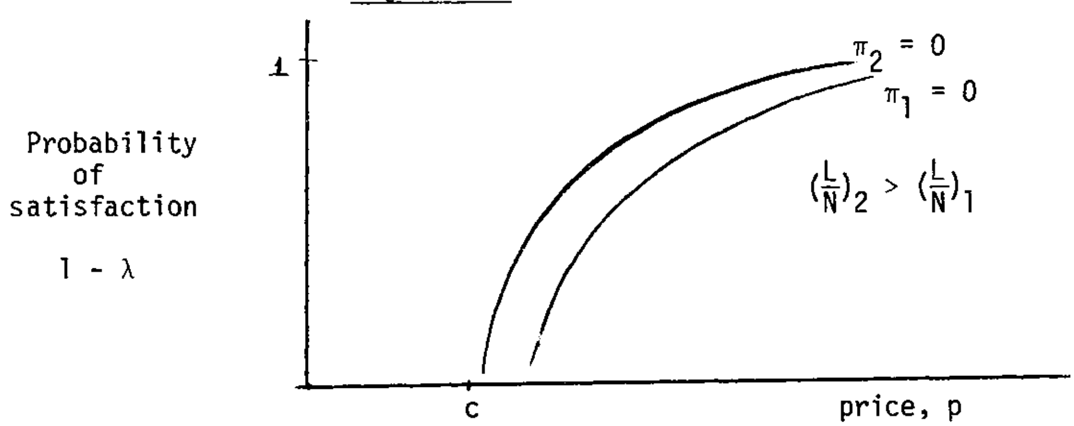
As the customer per firm ratio, L/N, increases, the $\pi = 0$ curve is affected in several ways. First, as already mentioned, the price along the horizontal axis at which the $\pi = 0$ curve begins to rise moves closer to the cost of production, c. Secondly, the entire curve shifts up, indicating that for fixed price as the number of customers per store increases firms can afford to increase their customer capacity in such a way that there is a higher probability of satisfying customers. Basically, this result occurs be-

[1]The probability that a firm obtains at least one customer is approximately $1 - e^{-L/N}$.

cause there are economies of scale in servicing a stochastic market. The proportional risk of having unsold goods declines as the customer per firm ratio increases. In other words, to achieve a satisfaction probability of .5 in a market with 100 customers per store requires a $\frac{s}{100}$[1] figure that is larger than the $\frac{s}{1000}$ figure in a market with 1000 customers per store. As the customer per firm ratio continues to increase, the $\pi = 0$ curve shifts up to the $1 - \lambda = 1$ line.

How does the slope of the $\pi(1 - \lambda, p) = 0$ curve behave as the customer per store ratio increases? It is possible to prove that for any fixed price, p, greater than c, the slope $(\frac{d(1-\lambda)}{dp})$ falls monotonically to zero as L/N increases. Furthermore, for any fixed probability of satisfaction, $\overline{1 - \lambda}$, below 1, the slope $\frac{d\ 1-\lambda}{dp} \mid \overline{1 - \lambda}$ approaches infinity as L/N increases. These properties are illustrated in the diagram below and are proved in the appendix to this chapter.

Figure 2-7 - The Zero Profit Curve and the Customer Per Firm Ratio



2.6.4 Behavior of Market Equilibrium as the Customer Per Firm Ratio Increases

Armed with these properties of the $\pi = 0$ curve, we can now investigate the behavior of equilibrium as the customer per firm, L/N, ratio increases.

[1]Recall that s refers to customer capacity.

It will be useful for the reader to recall from the discussion on consumer preferences that b and B are the lower and upper bounds on the slope of the isoutility curves, respectively.

Theorem 1: As the customer per firm ratio, L/N, increases, the equilibrium price associated with the market clearing point approaches the deterministic market clearing price c.

Proof: The method of proof will be to show that as L/N increases, the equilibrium point $(p^*, 1 - \lambda^*)$ of the market clearing under uncertainty will eventually lie to the left of the vertical line $p = c + e$ for every positive e.

Choose the point $p = c + e$ for any positive e. Choose L/N large enough so that the $\pi = 0$ curve is defined by (5) for p equal $c + e$. Equilibrium in the uncertain market is defined as the point of tangency between the $\pi = 0$ curve and the highest isoutility curve.[1] Now, increase L/N. As L/N increases, the slope of the $\pi = 0$ curve declines to zero for any fixed $p > c$. Increase L/N so that the slope of the $\pi = 0$ curve is less than b at $p = c + e$. This implies that the slope of $\pi = 0$ is less than b for all $p \geq c + e$, since the $\pi = 0$ curve is concave. But then it is impossible for any isoutility curve to be tangent to the $\pi = 0$ curve at any price above $c + e$. Hence, the market equilibrium price p^* is less than $c + e$. Since p^* must be greater than c for any production to occur at all, and since p^* is less than $c + e$ for any positive e, it follows that $\lim_{\frac{L}{N} \to \infty} p^* \to c$. Q.E.D.

[1]Recall that we are excluding the uninteresting case of boundary solutions. Actually, since for sufficiently large L/N the slope of the $\pi = 0$ curve is arbitrarily large for low prices, and arbitrarily small for high prices, the Lipschitz conditions on the isoutility curves rule out the possibility of boundary solutions.

<u>Theorem</u> 2: As the customer per firm ratio, L/N, increases, the equilibrium probability of satisfaction approaches 1.

<u>Proof</u>: The method of proof will be to show that as L/N increases, the equilibrium point (p^*, $1 - \lambda^*$) lies above the horizontal line defined by probability of satisfaction = $1 - \bar{\lambda}$ for $1 - \bar{\lambda} < 1$.

As before, equilibrium is determined by the point of tangency between the $\pi = 0$ curve and the highest isoutility curve. Choose any $1 - \bar{\lambda} < 1$. Increase L/N. As L/N increases, the slope along $\pi = 0$ curve at the point associated with a probability of satisfaction equal to $1 - \bar{\lambda}$ becomes arbitrarily large. Continue increasing L/N until the slope at $1 - \bar{\lambda}$ on the $\pi = 0$ curve exceeds B. Because of the concavity of the $\pi = 0$ curve, the the slope along the $\pi = 0$ curve exceeds B for all $1 - \lambda < 1 - \bar{\lambda}$. Hence, for sufficiently large L/N, it is impossible for any isoutility curve to be tangent to the $\pi = 0$ curve for a probability of satisfaction less than or equal to $1 - \bar{\lambda}$. Since the equilibrium probability of satisfaction is bounded above by 1, and lies above every $1 - \bar{\lambda}$ less than 1, it follows that $\lim_{\frac{L}{N} \to \infty} 1 - \lambda^* \to 1$. Q.E.D.

It immediately follows from Theorems 1 and 2 that the equilibrium level of expected utility achievable by consumers in equilibrium approaches the level of utility achievable in the deterministic market, where price equals c and the probability of satisfaction equals one.

<u>Theorem 3</u>: As the customer per firm ratio, L/N, increases, the percent discrepancy between the amount supplied and the amount demanded approaches zero.

Proof: The total amount demanded equals the number of customers times the per capita demand, $L \cdot x(p)$, while the total amount supplied equals the number of firms times the customer capacity per firm times the per capita demand, $N \cdot s \cdot x(p)$. To prove the Theorem it is sufficient to show that $\frac{N \cdot s}{L} \to 1$ as $\frac{L}{N}$ increases.

Using (2) and (4), we can write that the zero profit condition implies

$$(1 - \lambda)p \cdot L = Nc \cdot s$$

From the previous two theorems we know that in equilibrium $p \to c$ and $1 - \lambda \to 1$ as L/N increases. Hence the Theorem follows immediately. Q.E.D.

Theorem 3 dealt with the percent discrepancy between supply and demand. What about the absolute discrepancy, $[L - N \cdot s]x(p)$ - does that too approach zero as the customer per firm ratio, L/N, increases? The answer in general is no. Usually the absolute discrepancy will approach either plus or minus infinity as L/N increases. In other words, equilibrium is possible even though the number of dissatisfied customers is arbitrarily large.

To see how to construct an example where the discrepancy between supply and demand becomes arbitrarily large, we will use a result from the appendix to this chapter. In the appendix, we show that as L/N increases, the slope of the $\pi = 0$ curve at the equilibrium point equals

$$\frac{d}{dp}\frac{1-\lambda}{} = \frac{1}{c}\frac{1-F(u)}{F(u)} \tag{6}$$

where as before $u = \frac{s-\sigma^2}{\sigma}$, F = cumulative normal, and $\sigma^2 = \frac{L}{N}$.

Since $Ns = L + Nu\sigma$, we see that the absolute discrepancy between supply and demand will go to zero as the customer per firm ratio, L/N, increases only if u goes to 0. From (6), we see that as L/N increases, the slope of the zero profit curve at the equilibrium point that corresponds to a zero value for u, equals 1/c. Since equilibrium is determined by the tangency between the zero profit and highest isoutility curve, it is evident that as the equilibrium price - probability of satisfaction approaches (c, 1) as L/N increases, the value of u in equilibrium will equal zero if and only if the slope of the isoutility curves around the point (c, 1) is $\frac{1}{c}$ - otherwise the tangency will occur at a point that corresponds to a value of u other than zero.

If the slope of the isoutility curves is strictly greater than 1/c around the point (c,1), then u will be bounded away from zero, and will be negative. In such a case, the difference between supply and demand in the market equilibrium will be negative and grow indefinitely large as L/N increases. Demand will always exceed supply, yet there will be no incentives for firms to increase their supply. Equilibrium takes into account consumer preferences for both price and probability of satisfaction. Any firm that tried to increase its supply and raise its price to cover its costs would lose business. By similar reasoning, if the slope of the isoutility curves around the point (c,1) is strictly less than 1/c, then as L/N increases, equilibrium will involve supply exceeding demand by larger and larger amounts.

It is clear then that as L/N increases, equilibrium will involve total supply and demand being in balance only in one special case when the isoutility curves have a slope of $\frac{1}{c}$ around the point (c,1). In general, as L/N increases, equilibrium will involve arbitrarily large absolute discrepancies between supply and demand.

In the corresponding deterministic market, in equilibrium, supply equals demand, price equals c, and the probability of satisfaction equals 1. The preceding theorems and discussion have shown that the equilibrium for the markets under study approaches this deterministic equilibrium in some, though not all, respects as the customer per firm ratio increases.

The reason why the market equilibrium does not converge to the deterministic one in all respects as the customer per firm ratio L/N increases can be explained as follows. As L/N increases, the total uncertainty in the market increases, so that market operation under uncertainty differs considerably from that under certainty. On the other hand, by the law of large numbers, the proportional risks caused by the uncertainty vanish as L/N increases. Therefore, percentage-wise concepts (e.g. supply ÷ demand), or concepts that apply to individual units of the good (e.g. price) or individual customers (e.g. probability of satisfaction) approach their values in the corresponding deterministic market as L/N increases. However, aggregate concepts such as supply, demand, and total number of customers dissatisfied do not, in general, approach their values in the deterministic market as the customer per firm ratio increases.

2.6.5 Comparisons of Market Clearing Under Certainty and Under Uncertainty

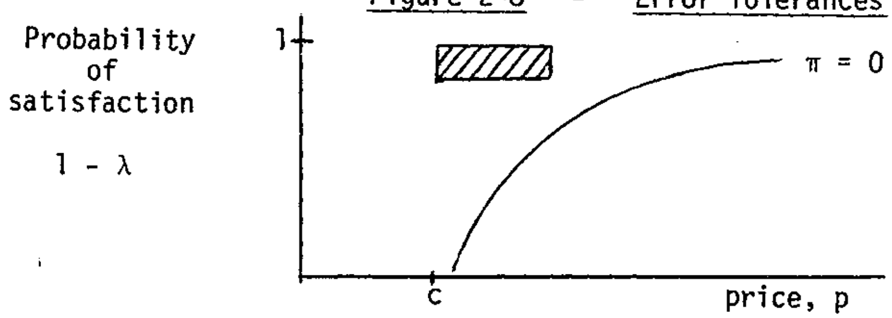
The reader might well be wondering just how important it is to examine market clearing under uncertainty by an analysis more complex than the simpler deterministic analysis that says price equals c, probability of satisfaction equals 1, and quantity supplied and demanded equals $L \cdot x(c)$. It is not possible to fully perceive the sharp differences between these uncertain markets and the traditional deterministic ones until the social welfare implications and especially the incentives facing interacting firms are examined. Still, at this stage, it is possible to give a preliminary evaluation.

First, for "moderate" values of the customer per firm ratio, L/N, it is evident that the deterministic analyses could lead one totally astray. As seen above, equilibrium will usually involve having supply and demand out of balance, a price in excess of c, and a probability of satisfaction below 1. Certainly, for moderate values for L/N, the deterministic analysis is simply inadequate.

What about for "large" L/N - can the deterministic analysis suffice there? As seen above, even as L/N increases, the discrepancy in equilibrium between supply and demand need not vanish and will in general become arbitrarily large. A deterministic analysis will completely miss this potentially important feature of market equilibrium. On the other hand, Theorems 1 and 2 do assure us that for "sufficiently large" values of L/N, the deterministic analysis will predict correctly the price and probability of satisfaction.

The question naturally arises as to how large does L/N have to be before the deterministic analysis is not too far wrong in its predictions of price and probability of shortage. To answer this question, the exact shape of the isoutility curves would be needed. Alternatively, we can ask the somewhat simpler (and less informative) question of what is the smallest value for the customer per firm ratio such that it is even possible for the deterministic analysis to be "approximately" correct. The value of L/N that answers this question will not tell us that for all larger values of L/N the deterministic analysis will suffice. Indeed, the value of L/N for which the deterministic analysis does suffice will usually considerably exceed the value of L/N that answers the preceeding question. What the answer to the question does tell us is that if L/N is less than the calculated L/N, the deterministic analysis will definitely fail. The calculated value for L/N provides a lower bound on the value of L/N that is required if the deterministic analysis is to even have a chance of satisfying the desired tolerance limits. In terms of the diagram below, if the shaded box represents acceptable errors for the deterministic analysis, we want to answer the question of how large L/N has to be before the zero profit curve hits the lower right hand corner of the shaded box. Call this value of the $\frac{L}{N}$ the "critical" L/N value.

Figure 2-8 - Error Tolerances



Only for values of L/N larger than this critical value can the equilibrium possibly occur in the shaded region. Of course, this critical value of L/N depends on the size of the shaded region which reflects the size of the allowable errors.

For example, if we are not very demanding and are willing to tolerate a 10% error in price (i.e. the actual equilibrium price $\leq$ 1.1.c) and a 10% error in the probability of shortage (i.e. actual equilibrium probability of satisfaction $\geq$ .90), then the deterministic analysis will have a chance of succeeding only if L/N exceeds 60.[1] If we tighten our tolerance limits to a 2% error in price and a 2% error in the probability of satisfaction, L/N must rise to 1600 before equilibrium could possibly fall in the shaded region. For more stringent requirements of only 1% errors in the price and probability of satisfaction, L/N must exceed 6500 before the deterministic analysis could even hope to meet the error standards.

[1] This value for L/N is calculated from (A23) of the appendix to this chapter.

Considering that these figures are lower limits on the value of L/N needed if the deterministic analysis is to suffice, it seems that for most purposes in order to be sure the deterministic analysis will not make large errors in the price and probability of shortage, we must require what for most markets is an uncomfortably large customer per firm ratio.

In summary, for moderate values of the customer per firm ratio, L/N, the deterministic analysis is inadequate. It is only for very large, perhaps unrealistically large, values for the customer per firm ratio that the deterministic analysis will be able to yield some useful results. Even then, however, the deterministic analysis will be unable to detect arbitrarily large absolute discrepancies between supply and demand. In Chapters 3, 4, and 5, it will be shown that the incentives for firms are much different in the uncertain environment under study than in the corresponding deterministic environment. The consequences of the different incentives will make even clearer the widespread differences between market clearing under certainty and under uncertainty.

2.7 More General Buying Behavior

2.7.1 Clumping

It is possible to introduce more general buying behavior into the model. For example, we could allow "clumping" of demand whereby $d(d > 1)$ customers, instead of one, show up at each visit to a firm (e.g. people shop with friends). If customers shop in clumps of size $d(d > 1)$, then in terms of the model it is equivalent to having per capita demand rise to $d \cdot x(p)$ and the total customer population fall to L/d. Basically, clumping causes the distribution of demand to look riskier [in the Rothschild-Stiglitz sense[1]] to the firm than it was in the case of no clumping. Since the height of the zero profit curve depends on the customer per firm ratio, clumping causes the zero profit curve to shift down from the no clumping ($d = 1$) case. Hence, when clumping occurs the equilibrium level of utility falls.

2.7.2 Random Per Capita Demand

The model has regarded the number of customers that a firm receives each period as random, but the per capita demand, $x(p)$, as deterministic. Allowing random demand per customer would not alter any of the important qualitative features of market operation that have already been discussed. Customers still have preferences for price and probability of shortage and firms still compete on the utility level they offer customers until profits are driven to zero.

The introduction of random independent per capita demand for customers would, however, make the analytics of the model intractable. If per capita demand is random, then the randomness must arise because of some unspecified stochastic components in the utility function. The probability of shortage will equal the expected value of unsatisfied demand divided by total demand. The dependence between the numerator and denominator will cause this probability of shortage to be difficult to derive. On the other hand, with some plausible rules of thumb, it turns out that the model with random per capita demand is analytically equivalent to the model that has been analyzed.

For example, at price p, consumers will achieve a level of utility $\tilde{u}(\tilde{x}(p), Y - \tilde{x}(p) \cdot p)$ where a "~" represents a random quantity and $\tilde{x}(p)$ is

[1] M. Rothschild and J. Stiglitz - "Increasing Risk I, A Definition" - Journal of Economic Theory, September, 1970.

the stochastic per capita demand such that $E(\tilde{x}(p)) = x(p)$, the deterministic demand of the previous model. When a consumer evaluates the prospect of buying at price p, he calculates the expected value of utility $\tilde{u}$, where expectations are taken over the unspecified random components in the utility function. Alternatively, we can postulate that consumers calculate their expected utility by substituting in the most likely point estimates of random variables - i.e., substitute $x(p)$ in place of $\tilde{x}(p)$ in the utility function. With this rule of thumb, the isoutility curves will be identical to those derived earlier.

Similarly, if we regard both the firm and the individual as calculating the probability of shortage as the expected shortage divided by the expected demand, we obtain the identical probability of satisfaction function derived earlier. In addition, it seems reasonable that a firm would calculate the amount it had to stock to satisfy approximately s customers as $s \cdot x(p)$. With these rules of thumb, the analytics of the model with random demands reduces to those derived earlier.

In summary, the introduction of random per capita demands does not alter any of the qualitative features of market operation, but can complicate the analytics of the model enormously. The reason for the complication is essentially that once individual demands are random, several of the quantities of interest involve expectations of quantities with random variables in both the numerator and denominator. However, under plausible rules of thumb for how a firm and a consumer cope with making decisions under uncertainty[1],

[1]Note, however, the asymetric treatment of uncertainty on the consumer side. The uncertainty introduced through the probability of satisfaction is treated explicitly while the uncertainty caused by the random and unspecified stochastic components in the utility function is treated by the plausible rules of thumb described above.

the case of the random per capita demand reduces to the model investigated in the previous sections of this chapter. The rule of thumb postulated to firms is most reasonable when the variation in the number of customers, and not the variation in per capita demand, is the main source of the firm's uncertainty about its demand. For many markets, it appears reasonable to assume that the major cause of uncertainty in demand to individual firms is not fluctuations in individual per capita demand, but rather fluctuations in the number of customers per period. Henceforth, whenever additional random features of market operation arise and their randomness is small compared to that of the number of customers per firm, we shall assume that the firms and the customers use rules of thumb, and behave as if these additional random variables took on their expected values.

2.8 Different Types of Customers

It is perfectly natural to imagine a market with two types of customers, who have different preferences between price and probability of satisfaction. In such a situation, it is possible to have an equilibrium in which two types of firms are established, each of which caters only to the preferences of one type of consumer. For example, suppose that there are two types of customers, and two types of firms. There are N_1 firms that cater to type 1 customers and N_2 ($N_2 < N_1$) firms that cater to type 2 customers. The equilibrium involving firm specialization is depicted in Figure 2-9.[1]

As Figure 2-10 illustrates, such specialized equilibrium may not always exist. The specialized equilibrium cannot exist because all the type 2 customers are better off at type 1 stores than at type 2 stores.

[1] For the case of equilibrium involving firm specialization, an outside observer might incorrectly conclude that there was a distribution of prices for an identical good, and attribute it to consumer ignorance. As this chapter emphasizes, since each type of firm offers a different probability of satisfaction, the "goods" at different types of firms are not identical.

Figure 2-9 - Specialized Equilibrium

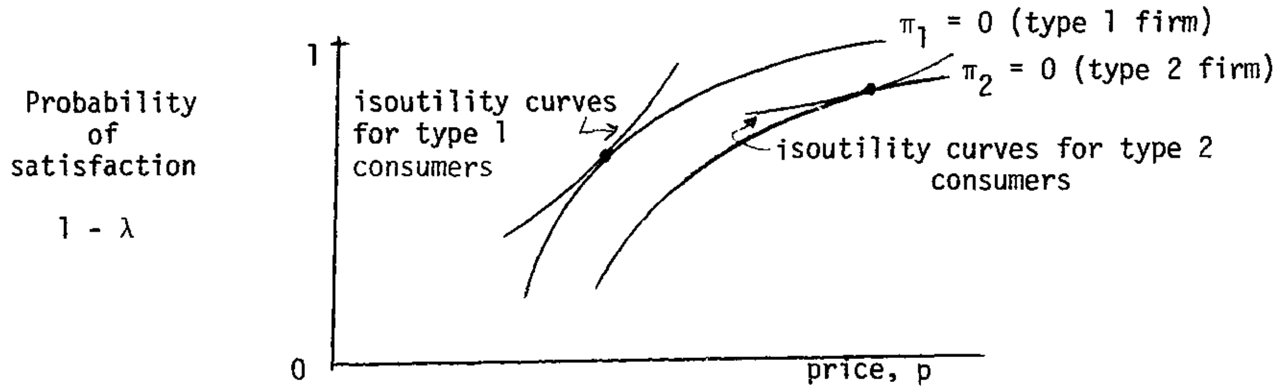
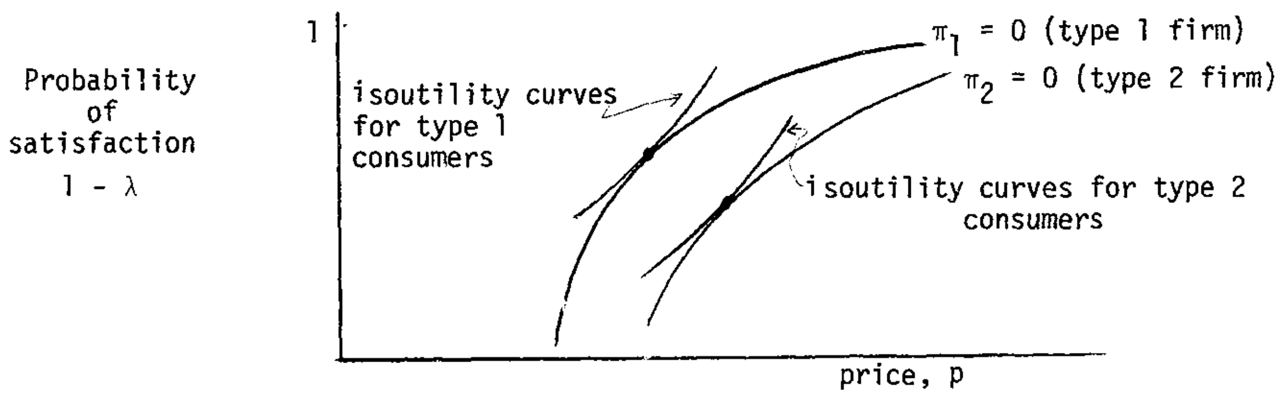


Figure 2-10 - Specialized Equilibrium Not Existing



When only one equilibrium can exist in the market, the question as to where it is established will be determined by the tastes of the majority. If any firm does not cater to the tastes of the majority, it will lose a majority of its business and will have to specialize in the minority's tastes. But, by assumption, specialized equilibriums are impossible, so the firm could not profitably attract just the minority types to its firm.

When the different consumers not only have different preferences, but also have different random patterns of demand, then the analysis becomes more complicated. Since randomness imposes costs on firms, if there is a nonspecialized equilibrium, then there will be an externality involved with low risk customers being forced to pay for the costs that high risk customers impose. As mentioned in the previous section, for a large number of markets, we expect the randomness of per capita demand to be much less important than the randomness in the number of customers. In such cases, the implications of random per capita demand can be ignored.

In cases where randomness of per capita demand is the central feature of market operation, the rules of thumb of the last section will not capture the implications on market operation of having customers who differ in their randomness of demand. The prime example of a market where differentiation by risk class is a crucial consideration is an insurance market. Studying such markets and the effects on equilibrium of having different risk classes of customers is beyond the scope of this research.[1]

2.9 Inventories and Search Behavior

The model assumes that firms cannot stock inventories of the good, and that consumers do not search at other firms if consumers are initially unable to obtain the good. Neither of these simplifying assumptions affects the qualitative behavior of market operation. As long as there are costs to holding inventory, a firm prefers not to be left with unsold goods at the end of the day. Therefore, just as in the simpler model, a firm is concerned with the risk of overstocking. Similarly, as long as there are costs to search, consumers prefer not going to firms which have high probabilities of being sold out of the good. As in the simpler model, equilibrium must take into account consumer's preferences for price and probability of shortage, and firms' ability to profitably operate at various price-probability of shortage combinations. Adding inventory holdings and

[1] For an analysis of insurance markets in which the different risk classes of consumers is the focus of study, see the work by Charles Wilson, especially "A Model of Insurance Markets", mimeo.

search behavior to the model would complicate the model but would add little insight into market operation.

2.10 The Number of Firms

In the model, the number of firms, N, is exogenous and is taken to be greater than one. Although this is a perfectly reasonable assumption to make in order to investigate the short run equilibrium for the markets under study, the question arises as to whether in the long run we expect the number of firms to exceed one, or whether we expect all N firms to merge into one giant firm. If two firms merge and are able to pool together their demand and their stocks, then the combined cost of operation of the merged firms will be lower than that of the unmerged firms. Basically, because of the stochastic environment, there are economies of scale as demand increases. If all N firms merge so that the number of firms equals one, then all uncertainty disappears since all customers would frequent the same firm. At first glance, it does appear that firms will have an incentive to merge into one giant firm. In this section, we suggest why the N firms may not have an incentive to merge together, and therefore why it is reasonable to expect there to be more than one firm in long run equilibrium for the markets under study.

First, there might be congestions costs associated with an N firm merger. In other words, there may be some increasing costs associated with horizontal merger and expansion that are not in the model. Transaction costs might overwhelm any gain from merger and thereby prevent one firm from desiring to merge with other firms.

A second and very important reason has to do with spatial location. When we are talking about merger, we imagine two or more of the firms combining operations at the same location. However, if demand is assigned randomly on a geographic basis, then if two firms merge at the same location their combined total demand could fall. If two firms merge but maintain separate locations, then there are no gains to mergers unless the merged firms can ship goods back and forth amongst themselves. But, in the model, the reason why it was assumed that firms cannot receive delivery of the good during the market period was presumably because such delivery was costly and/or time consuming. In such cases, there may well be no gains to spatially separate mergers.

It is clear then that there are good reasons to expect that for the markets under study the number of firms will exceed one in the long run equilibrium. For the remainder of this research, we shall usually regard the number of firms as fixed and greater than one, and not distinguish between short and long run concepts of equilibrium.

2.11 Summary

This chapter has presented a model of a market characterized by demand uncertainty, price inflexibility and a lead time in production. Prices are inflexible for a period over which demand is random. Competition among firms will lead to the establishment of a market equilibrium. Consumer preferences between the price and the risk of being unable to purchase the good determine equilibrium. In equilibrium, supply and demand will not in general be in balance. It is a natural feature of the equilibrium for the markets under study that there will be some customers unable to purchase the good at the same time that there are firms unable to sell their entire stock. As the customer per firm ratio increases, some components of the market equilibrium approach their values in the corresponding deterministic market. However, even for the determination of these components of equilibrium, the value of the customer per firm ratio that justifies using the deterministic analysis as a good approximation to the more complicated analysis seems unrealistically large. The absolute discrepancy between supply and demand will usually grow arbitrarily large as the customer per firm ratio increases. The qualitative features of the model would not be changed if the complications of inventory holdings and consumer search behavior were introduced, although the analytics of the model would become intractable.

The analysis of this chapter provides the basic groundwork to examine the welfare implications of these uncertain markets, and firm interaction in such markets. Although the differences between market clearing under uncertainty and under certainty are already evident, these differences will become even sharper in later chapters when issues of social welfare and firm interaction and examined.

Appendix A

Mathematical Appendix for Chapter 2

This mathematical appendix derives the geometric properties of the curves used in Chapter 2 that govern firm behavior. We will make extensive use of continuous normal approximations to the discrete binomial process. In doing this we are making a slight approximation error, since the normal process allows certain inherently positive quantities to become negative (though with a very small probability). For moderate values for the customer per store ratio (i.e. $L/N \simeq 40$), the continuous approximation will be very good. In order to simplify the analytics, we will assume that the customer capacity, s, is always larger than some small number, say 5. Since all the theorems of Chapter 2 have the customer per store ratio, L/N, increasing to infinity, we see that this assumption is very weak. Allowing the minimum s to go to zero would greatly complicate the proofs, but would not add any insights into how markets operate under uncertainty.

Consider the binomial process of assigning customers to any one firm. Let $\bar{s} = \frac{L}{N}$, which is the mean of the binomial process, and let $\sigma = \sqrt{\frac{L}{N}}$, which is the approximate variance of this binomial process. Now consider the expression for zero profits. All notation was previously defined in Chapter 2. From Chapter 2, we have that

$$\pi = [p \sum_{0}^{s} i\ pr(i) + ps \sum_{s+1}^{\infty} pr(i) - cs]x(p) = 0 \quad ,$$

$$\text{or } \pi = p\sigma \sum_{0}^{s} \frac{(i-\bar{s})}{\sigma} pr(i) + p\bar{s} \sum_{0}^{s} pr(i) + ps[1 - \sum_{0}^{s} pr(i)] - cs = 0, \text{ [1]}$$

where pr(i) is the binomial probability of obtaining i customers.

---

1 We rule out the case of x(p) = 0 as uninteresting. Notice that the per capita demand x(p) will not affect the zero profit curve.

From the Central Limit Theorem, we know that $\frac{i-\bar{s}}{\sigma}$ is distributed approximately as N(0,1). Rewrite the above expression as

$$\pi = p\sigma N^{E\ell}(u) + p\bar{s}\ F(u) + ps[1 - F(u)] - cs = 0,$$

where

$N^{E\ell}(u) = \int_{-\infty}^{u} tf(t)dt$, f = normal density, F(u) = cumulative normal distribution, and $u = \frac{i-\bar{s}}{\sigma}$.

Define $I(u) = N^{E\ell}(u) - uF(u) = \int_{-\infty}^{u}(t-u)f(t)dt$. From its definition, it is easy to establish that the function I(u) has the following properties:

1) $I(u) < 0$

2) $I'(u) = -F(u)$

3) $\lim_{u \to -\infty} I(u) = 0$

4) $\lim_{u \to \infty} I(u) = -\infty$.

We can rewrite the expression for profits equal to zero as

$$p\sigma\ N^{E\ell}(u) - p\ (\frac{s-\bar{s}}{\sigma})\ F(u) + s[p - c] = 0\ , \text{ or}$$

$$p\sigma[N^{E\ell}(u) - uF(u)] + s[p - c] = 0\ , \text{ or}$$

$$p[\sigma I + s] - sc = 0 \qquad \text{(A1), or}$$

along the zero profit curve,

$$p = cs/\sigma I + s \quad . \qquad \text{(A2)}$$

From (A1), it immediately follows that the continuous approximation to the binomial is valid only for $\sigma I + s > 0$. Moreover, after some manipulation, the zero profit (henceforth $\pi = 0$) equation, can be written as

$$p[N^{E\ell}(u) + p\bar{s}F(u)] + [p[1 - F(u)] - c]s = 0 \quad . \qquad (A3)$$

Since $0 < \sum_{0}^{s} i\, pr(i) \approx \sigma N^{E\ell}(u) + \bar{s}F(u)$, it follows from (A3) that the normal approximation is valid only when $p - c - pF(u) < 0$.

Implicitly differentiating (A1) with respect to p, we obtain

$$[\sigma I + s] + p[-F(u) + 1]\frac{ds}{dp} - \frac{ds}{dp}c = 0$$

or

$$\frac{ds}{dp} = \frac{[\sigma I + s]}{[p - c - pF(u)]}(-1) \quad . \qquad (A4)$$

From the remarks of the previous paragraph, we see that $\frac{ds}{dp} > 0$.

If we differentiate (A2), we obtain an alternative expression form $\frac{ds}{dp}$, namely

$$\frac{ds}{dp} = (\sigma I + s)\frac{(\sigma I + s)}{(\sigma I + sF)c} \quad . \qquad (A5)$$

Equating (A4) to (A5), it follows that along the $\pi = 0$ curve,

$$\frac{(\sigma I + s)}{(\sigma I + sF)}\frac{1}{c} = \frac{-1}{p - c - pF} \qquad (A6)$$

and hence it follows that $\sigma I + sF > 0$ over the relevant range.

We are interested in determining the properties of the $\pi = 0$ curven when drawn in $(1-\lambda,\ p)$ space. Let us derive what the probability of shortage, $1-\lambda$, looks like as a function of s. The expected shortage, M, for any one store with customer capacity s is given by

$$M(s) = \sum_{s+1}^{\infty} (i - s)\ pr(i) \quad , \quad \text{or}$$

$$M(s) = \sum_{s+1}^{\infty} (i - \bar{s})\ pr(i) + (\bar{s} - s)\ pr(i) \quad , \text{or}$$

$$M(s) = -\ \sigma N^{E\ell}(u) - u\sigma(1 - F(u)) \quad ,$$

where as before $u = \frac{s - \bar{s}}{\sigma}$. If all N stores follow the same operating policy, then the total shortage is $N \cdot M(s)$. The total fraction of dissatisfied customers is then

$$\lambda(s) = \frac{N \cdot M(s)}{L} \tag{A7}$$

or

$$\lambda(s) = \frac{M(s)}{\sigma^2} \quad .$$

Hence, $1 - \lambda(s) = 1 - \frac{M(s)}{\sigma^2}$, or

$$1 - \lambda(s) = \frac{\sigma^2 + \sigma N^{E\ell} + \sigma u[1 - F(u)]}{\sigma^2}$$

$$\text{or} \qquad 1 - \lambda(s) = \frac{\sigma[N^{E\ell}(u) - uF(u)] + (\sigma^2 + u\ )}{\sigma^2}$$

$$\text{or,}\ 1 - \lambda(s) = \frac{\sigma I + s}{\sigma^2} \ . \qquad \text{(A8)}$$

Differentiating (A8), we find that

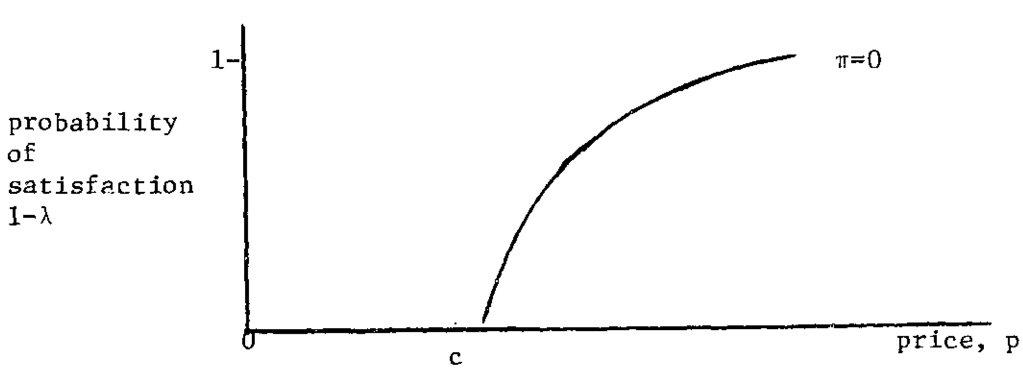
$$\frac{d[1 - \lambda(s)]}{ds} = \frac{1 - F(u)}{\sigma^2} > 0 \ . \qquad \text{(A9)}$$

From (A9) and the remark below (A4) it follows that

$$\left.\frac{d(1 - \lambda)}{dp}\right|_{\pi=0} = \frac{d(1 - \lambda(s))}{ds} \frac{ds}{dp} > 0 \ . \qquad \text{(A10)}$$

In $(1 - \lambda, p)$ space, (A10) tells us that the $\pi = 0$ curve slopes upwards, as depicted below.

Figure A1 -- The $\pi = 0$ Curve



Notice that the $\pi = 0$ curve begins a little to the right of the point $p = c$. The curve always lies to the right of the line $p = c$, because the price always has to be a little above the production cost c to cover the cost of unsold goods. The next two propositions provide information on how the slope $\frac{d\,1-\lambda}{dp}$ behaves for very low and very high prices. For these next two propositions, we ignore the technicality that the range for the number of customers is not $(-\infty,\infty)$ as the continuous approximation treats it, but really (0,L). This technicality is of minor importance below since the continuous approximation will assign an extremely small probability to the nonfeasible range.

Proposition 1: The slope of the $\pi = 0$ curve falls to zero as price increases.

Proof: From (A5), (A7) and (A10), we have

$$\frac{d(1-\lambda)}{dp} = \frac{1}{\sigma^2}\,\frac{(1-F)}{c}\,\frac{(\sigma I + s)^2}{\sigma I + sF} .$$

Using (A8), we have

$$\frac{d(1-\lambda)}{dp} = \frac{(1-F)}{c}\,(1-\lambda)\,\frac{(\sigma I + s)}{(\sigma I + sF)} .$$

From the definition of profits it is immediately obvious that as $p \to \infty$, $s \to \infty$, which implies $u \to \infty$. Clearly, as $s \to \infty$, $1 - \lambda \to 1$, and $\frac{\sigma I + s}{\sigma I + sF} \to 0$, and $1 - F \to 0$. Hence, the result follows. Q.E.D.

Proposition 2: The slope of the $\pi = 0$ curve approaches infinity as the price falls to c (more precisely, as the price falls to the value at which the $\pi = 0$ curve begins).

Proof: From the definition of profits, it follows that as $p \to c$, that $s \to -\infty$ (actually 0), and $u \to -\infty$. Since

$$\frac{d(1-\lambda)}{dp} = \frac{1}{\sigma^2}\,\frac{1}{c}\,\frac{(\sigma I + s)^2}{\sigma I + sF}\,(1 - F),$$

we have that, provided the limit exists,

$$\lim_{s \to -\infty} \frac{d(1-\lambda)}{dp} = \frac{1}{c\sigma^2} \cdot \lim_{s \to -\infty} \frac{(\sigma I + s)^2}{\sigma I + sF}$$

since $\lim_{s \to -\infty} 1 - F \to 1$. By L'Hopital's rule and providing the limits exist we have that

$$\lim_{s \to -\infty} \frac{d(1-\lambda)}{dp} = \lim \frac{2}{c\sigma^2}\,\frac{[\sigma I + s]}{sf\,\frac{1}{\sigma}}\,(1 - F) = \lim \frac{2}{c\sigma}\,\left[\frac{\sigma I + s}{sf}\right].$$

By L'Hopital's rule, $\lim_{s \to -\infty} \frac{\sigma I + s}{sf} = \lim_{u \to -\infty} \frac{1 - F}{f + s\frac{(-u)}{\sigma}f} \to \infty$.

Hence, $\lim_{s \to -\infty} \frac{d(1-\lambda)}{dp} \to \infty$. Q.E.D.

The above results illustrate that the $\pi = 0$ curve starts off with a very steep slope near $p = c$, rises, and then approaches the $1-\lambda = 1$ line with a very flat slope. The previous two propositions were intended to give the reader some idea of what the $\pi = 0$ curve looked like. For that reason, we did not pay much attention to the details about not allowing variables to exceed their natural bounds. The next result establishes the concavity of the $\pi = 0$ curve. Since this theorem is of importance in Chapter 2, we pay careful attention to the range of values that variables take on.

Before proving the concavity of the $\pi = 0$ curve, we establish two useful lemmas:

<u>Lemma 1</u> Define $T_1(u) = \frac{2(1 - F)}{f} - \frac{s}{\sigma} = \frac{2(1 - F)}{f} - \sigma - u$.

For $u \geq 0$, $T_1(u) \leq 0$ for $\sigma > 3$.

<u>Proof:</u> Taking derivatives of $T_1(u)$, we obtain

$$T_1'(u) = 2[-1 + \frac{(1 - F)u}{f}] - 1 .$$

Since $\frac{(1 - F)u}{f} < 1$ for $u \geq 0$ [(1)] , it follows that

$T_1'(u) < 0$, and hence

$$\max_{u>0} T_1(u) = T_1(0) \doteq 2.5 - \sigma$$

Hence, for $\sigma > 3$, $T_1(u) < 0$ $\quad u \geq 0$. Q.E.D.

---

1 This inequality follows from Birnbaum's inequality which states that $\frac{1 - F(u)}{f(u)} < \frac{4}{3u + \sqrt{u^2 + 8}}$, $u > 0$. M. Kendall and G. Stuart, <u>The Advanced Theory of Statistics</u>, Vol. 1, p. 146.

<u>Lemma 2</u> Define $T_2(u) = \frac{-4F}{f} + \frac{s}{\sigma} = \frac{-4F}{f} + \sigma + u$. For $u \leq 0$,

$T_2(u) \geq 0$, for $\sigma > 6$.

<u>Proof:</u> Taking derivatives of $T_2(u)$, we obtain

$$T_2'(u) = -4[\frac{f}{f} + \frac{(-1)Fu f(-1)}{f^2}] + 1, \quad \text{or}$$

$$T_2'(u) = -4[1 + \frac{uF}{f}] + 1 \quad .$$

Since $\frac{-uF}{f} \simeq 1 - \frac{1}{u^2}$ for $u < -1$,[1] we have that

$T_2'(u) \approx - 4 \frac{1}{u^2} + 1$, so that for $u < - 2$, $T_2'(u) > 0$.

Therefore min $T_2(u)$ occurs at the u that corresponds to the minimum s of 5. If $s_{min} = 5$, then $u_{min} = -\sigma + \frac{s}{\sigma}$. Hence,

$\min_{u<-2} T_2(u) = T_2(u_{min}) = -4 \frac{F}{f} + \sigma + u_{min} \simeq -4 \frac{1}{(-u)} + \sigma + u_{min}$, since

$\frac{F(u_{min})}{f(u_{min})} \approx (\frac{1}{(-u_{min})})$. Rewriting the above, we have

---

1 See M. Kendall and A. Stuart, <u>op. cit.</u>, p. 137.

$$\min_{u_{min}<u<-2} T_2(u) \approx \frac{-4}{\sigma - \frac{5}{\sigma}} + \sigma + (-\sigma) + \frac{5}{\sigma}$$

$$= \frac{1}{\sigma - \frac{5}{\sigma}} \left[-4 + 5 - \frac{25}{\sigma^2}\right]$$

$$= \frac{\sigma}{\sigma^2 - 5} \left[1 - 25/\sigma^2\right]$$

$$= \frac{1}{\sigma} \frac{\sigma^2 - 25}{\sigma^2 - 5}, \quad \text{or}$$

$$\min_{u_{min}<u<-2} T_2(u) > 0 \qquad \text{for } \sigma > 5 .$$

For $u \varepsilon [-2, 0]$ we have

$T_2(u) = -4 \frac{F}{f} + \sigma + u$ or by Birnbaum's Inequality

$$\min T_2(u) > \frac{-4.4}{3|u| + \sqrt{u^2 + 8}} + \sigma + u$$

from which it follows that for $u \varepsilon [-2, 0]$,

$$\min T_2(u) > \sigma - \frac{16}{\sqrt{8}} = \sigma - 5.6 \text{ , or}$$

$$\min T_2(u) > 0 \quad \text{for } \sigma > 6 .$$

Hence it follows that $T_2(u) \geq 0$, $\forall u \leq 0$, if $\sigma > 6$. Q.E.D.

Using Lemmas 1 and 2, we can now prove the following.

<u>Proposition 3</u> For $\sigma > 6$, the $\pi = 0$ curve is concave in p -- i.e. $\frac{d^2(1 - \lambda)}{dp^2} < 0$ along the $\pi = 0$ curve.

<u>Proof:</u> We know from (A5), (A9) and (A10) that

$$\frac{d\,(1-\lambda)}{dp} = \frac{1}{\sigma^2}\,\frac{[1-F]}{c}\,\frac{[\sigma I + s]^2}{[\sigma I + sF]} \quad , \qquad \text{(A11)}$$

Since $\frac{d^2(1-\lambda)}{dp^2} = \frac{d}{ds}\, d(1-\lambda)/dp \cdot \frac{ds}{dp}$, and $\frac{ds}{dp} > 0$

(see (A4)), it is sufficient to prove that

$$\frac{d[d(1-\lambda)/dp]}{ds} < 0 \ .$$

There are two cases to consider, $u \geq 0$, and $u_{min} < u < 0$. First consider the case where u is positive.

Differentiating the expression (A11) we obtain

$$\frac{d(d1-\lambda/dp)}{ds} = \sigma^2 \frac{1}{c}\left(\left[-\frac{f}{\sigma}\,\frac{[\sigma I + s]^2}{[\sigma I + sF]}\right]\right.$$
$$\left. + [1-F]\left[\frac{2[\sigma I + s][1-F]}{\sigma I + sF} - \frac{[\sigma I + s]^2 sf/\sigma}{[\sigma I + sF]^2}\right]\right) \ ,$$

To establish that the above expression is negative it is sufficient to prove that the expression in the second set of brackets is negative. The sign of this term is the same as

$$\left[2(1-F) - \frac{\sigma I + s}{\sigma I + sF}\,\frac{sf}{\sigma}\right] \quad . \qquad \text{(A12)}$$

This expression is less than

$$2[1 - F] - \frac{sf}{\sigma}\ , \text{ since } \frac{\sigma I + s}{\sigma I + sF} > 1\ .$$

$$\text{Now sgn } \frac{d^2(1-\lambda)}{dp^2} = \text{sgn } [2(1 - F) - \frac{sf}{\sigma}] = \text{sgn}[2\frac{1 - F}{f} - \frac{s}{\sigma}] =$$

$$\text{sgn } T_2(u)$$

which is negative by Corollary 1 for $u \geq 0$ and $\sigma > 6$.

Now, consider the case where $u \leq 0$. We will use the relation (A6) that

$$\frac{\sigma I + s}{\sigma I + sF} = \frac{-c}{p - pF - c}\ .$$

Rewrite (A12) as

$$[2(1 - F) + \frac{c}{p - pF - c}\ \frac{sf}{\sigma}\ ]$$

$$= \frac{1}{p - pF - c}\ [2(1 - F) \cdot [p(1 - F) - c] + \frac{csf}{\sigma}\ ]\ .$$

We wish to show that the expression in brackets is always positive since $p - pF - c < 0$. Rewrite the expression in brackets as

$$2p(1 - F)^2 - 2(1 - F)c + c\frac{sf}{\sigma} = 2p - 4pF + 2pF^2$$

$$- 2c + 2Fc + c\frac{sf}{\sigma}\ ,$$

or

$$2[p - c] - 4pF + 2Fc + 2pF^2 + c\,\frac{sf}{\sigma} \quad .$$

The relation $p - pF - c < 0$ implies $c > p - pF$ so

$$2[p - c] - 4pF + 2Fc + 2pF^2 + \frac{csf}{\sigma} > 2[p - c] - 4pF$$

$$+ 2pF(1 - F) + 2pF^2 + \frac{p(1 - F)fs}{\sigma}$$

$$= 2[p - c] - 2pF + p(1 - F)\,\frac{fs}{\sigma} \, ,$$

$$= 2[p - c] + p[-2 + p(1 - F)\,\frac{fs}{\sigma}\,] \, .$$

For $u \leq 0$, $1 - F \geq \frac{1}{2}$ so

$$2[p - c] + p[-2F + (1 - F)\,\frac{fs}{\sigma}] > 2[p - c] + p[-2F + \frac{1}{2}\,\frac{fs}{\sigma}]$$

$$= 2[p - c] + \frac{p}{2F}\,[\,-\frac{F}{f} + \frac{s}{\sigma}]$$

$$= 2[p - c] + \frac{p}{f}\,[T_2(u)] \quad .$$

Since $p > c$ and $T_2(u) > 0$ for $u \leq 0$ and $\sigma > 6$ by the previous lemma, we have established the desired result. Hence $\frac{d^2(1 - \lambda)}{dp} < 0$ and the $\pi = 0$ curve is concave. Q.E.D.

The above proof required that the customer capacity s exceeded 5. All the results of Chapter 2 which depend on the concavity of the $\pi = 0$ curve deal with the case where the customer per store ratio, L/N, goes to infinity. Notice that for $s > 5$, that $1 - \lambda$ approaches 0 as L/N increases, so that this requirement of $s > 5$ becomes an extremely weak, though analytically convenient, assumption. We now want to investigate how the $\pi = 0$ curve behaves as σ (or L/N) increases.

<u>Proposition 4</u> As σ (or L/N) increases, the $\pi = 0$ curve shifts up in $(1 - \lambda, p)$ space (i.e. $\frac{d\,1-\lambda}{d\sigma} > 0$).

<u>Proof:</u> First, we need to calculate $\frac{ds}{d\sigma}$ from the $\pi = 0$ relation (A1)

$$[\sigma I] + [1 - c/p]s = 0.$$

Differentiating wrt σ, we obtain

$$I + \sigma I'(u)\ \left[\frac{\partial u}{\partial \sigma} + \frac{\partial u}{\partial s}\frac{ds}{d\sigma}\right] + [1 - c/p]\frac{ds}{d\sigma} = 0, \text{ or}$$

$$I - F[(-1)[\frac{s}{\sigma^2} + 1] + \frac{1}{\sigma}\frac{ds}{d\sigma}] + [1 - c/p]\frac{ds}{d\sigma} = 0 \text{ , or}$$

$$I + F[\frac{s}{\sigma} + \sigma] - F\frac{ds}{d\sigma} + [1 - c/p]\ ds/d\sigma = 0 \text{ , or}$$

$$\frac{ds}{d\sigma} = (-1)\ [I + \frac{sF}{\sigma} + \sigma F]/(1 - c/p - F) . \tag{A13}$$

Since $1 - c/p - F < 0$, and $I = \frac{sF}{\sigma} > 0$, it follows the $\frac{ds}{d\sigma} > 0$.

Let us now use the relation along the $\pi = 0$ curve that

$$1 - \lambda = \frac{1}{\sigma^2}\, s\, \frac{c}{p} \text{ , so that}$$

$$\frac{d(1-\lambda)}{d\sigma} = \frac{c}{p}\ [(-1)\frac{2s}{\sigma}\cdot\frac{1}{\sigma^2} + \frac{1}{\sigma^2}\frac{ds}{d\sigma}] = \text{sgn}\ [\frac{-2s}{\sigma} + \frac{ds}{d\sigma}] \text{ , or}$$

$$= \text{sgn}\ [\frac{-2s}{\sigma} + (-1)\ \frac{[I + \frac{sF}{\sigma} + \sigma F]}{1 - c/p - F}\ ] \ ,$$

$$= \text{sgn}\ [\frac{-2s}{\sigma} + (-1)\ \frac{\frac{s}{\sigma}\ [\frac{\sigma I}{s} + F + \frac{\sigma^2}{s}F]}{(1 - c/p - F)}\ ] \text{ , or}$$

using that $1 - c/p = -\sigma I/s$, $= \text{sgn}[\frac{-2s}{\sigma} + (-1)\frac{s}{\sigma}\frac{[(-1)[1-c/p]+F+\frac{\sigma^2}{s}F]}{(1-c/p-F)}$, or

$$= \text{sgn}[\frac{(-2s)}{\sigma} + \frac{s}{\sigma}) - \frac{\sigma F}{(1 - c/p - F)}] \text{ , or}$$

$$= -\text{ sgn } [-\frac{s}{\sigma}[1 - c/p] + \frac{s}{\sigma} F - \sigma F] \text{ , or}$$

since $-[1- c/p]\frac{s}{\sigma} = I$, and $\frac{s}{\sigma} = u + \sigma$

$$= -\text{ sgn } [I + (u + \sigma)F - \sigma F]$$

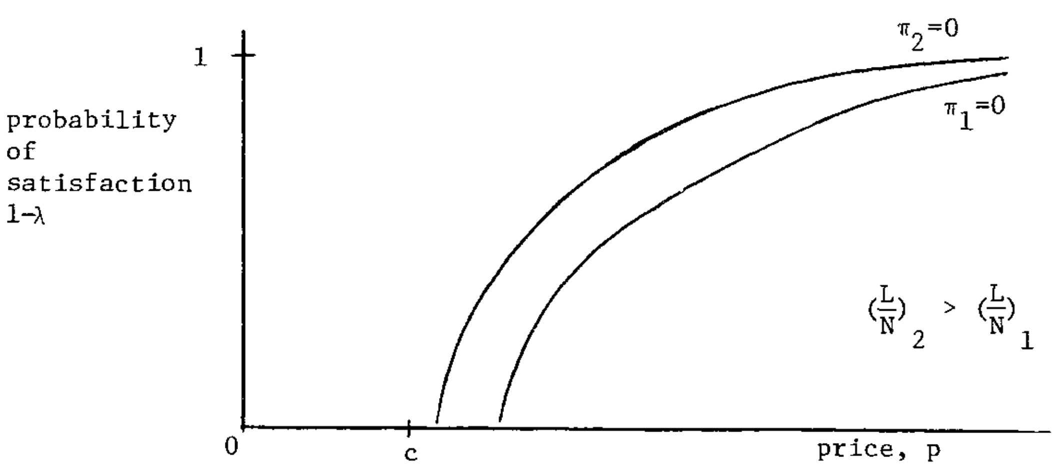
$$= -\text{sgn}[I + uF] \text{ , or}$$

since $I(u) = -f(u) - uF(u)$,

$$= -\text{ sgn}[-f(u)] = \text{sgn}[f(u)] > 0 \text{ .} \qquad \text{Q.E.D.}$$

Proposition 4 is illustrated below.

Figure A2 -- The Zero Profit Curve and the Customer Per Firm Ratio



We now want to show that for fixed p, the $\pi = 0$ curve shifts up to $1 - \lambda = 1$ as σ (or L/N) increases. We need the following lemma.

<u>Lemma 3</u> Along the $\pi = 0$ curve, for fixed p, as σ increases, u approaches infinity and $\frac{u}{\sigma}$ approaches $\frac{p}{c} - 1$.

<u>Proof:</u> Since $u \equiv \frac{s - \sigma^2}{\sigma} = \frac{s}{\sigma} - \sigma$, we have that

$$\frac{du}{d\sigma} = -1 - \frac{s}{\sigma^2} + \frac{1}{\sigma}\frac{ds}{d\sigma} .$$

From (A6) and from (A13), write

$$\frac{ds}{d\sigma} = \frac{s}{\sigma} - \frac{\sigma F}{1 - c/p - F}$$, so that

$$\frac{du}{d\sigma} = -1 - \frac{s}{\sigma^2} + \frac{s}{\sigma^2} - F/(1 - c/p - F)$$, or

$$\frac{du}{d\sigma} = -1 - F/(1- c/p - F)$$, or

$$\frac{du}{d\sigma} = [-1 + c/p]/(1 - c/p - F) > 0$$, since

$c/p < 1$ and $1 - c/p - F < 0$. Furthermore,

since $|1 - c/p - F| < \infty$, it follows that

$\frac{du}{d\sigma} > m$ for some $m > 0$, and hence $u \to \infty$ as σ increases.

Consider $\lim_{\sigma\to\infty} \frac{u}{\sigma}$ along the $\pi = 0$ curve for fixed p. From above we have $\lim \frac{u}{\sigma} = \frac{}{\infty}$. Applying L'Hopital's rule we find that

$$\lim_{\sigma\to\infty} \frac{u}{\sigma} = \lim_{\sigma\to\infty} \frac{du}{d\sigma} = \frac{-1 + c/p}{-c/p} = p/c - 1. \qquad \text{Q.E.D.}$$

Proposition 5: As σ (or L/N) increases, the $\pi = 0$ curve shifts up to $1 - \lambda = 1$ for any fixed p.

Proof: From (A1) and (A8) we have that

$$(1 - \lambda) = \frac{c}{p}\frac{s}{\sigma^2} \quad , \text{ or}$$

$$1 - \lambda = \frac{c}{p}\left(1 + \frac{u}{\sigma}\right) .$$

From the preceding lemma $\frac{u}{\sigma} \to \frac{p}{c} - 1$ as σ increases. Hence $1 - \lambda \to 1$ as σ increases. Q.E.D.

The remaining two propositions have to do with the behavior of the slope $\frac{d(1 - \lambda)}{dp}$ of the $\pi = 0$ curve as σ (or L/N) gets large.

Proposition 6: For any fixed p, the slope of the $\pi = 0$ curve falls to zero as σ (or L/N) increase -- i.e. $\frac{d(1 - \lambda)}{dp} \to 0$ as $\sigma \to \infty$.

Proof: From (A11), we have that

$$\frac{d(1-\lambda)}{dp} = \frac{1}{c}\,\frac{1}{\sigma^2}\,(1-F)\,\frac{(\sigma I + s)}{\sigma I + sF} ,$$

Along the $\pi = 0$ curve, $\sigma I + s = cs/p$ and $\sigma I = [c/p - 1]s$, so we can rewrite the above as

$$\frac{d(1-\lambda)}{dp} = \frac{1}{c\sigma^2}\,(1-F)\,\frac{c^2s^2}{p^2}\,\frac{1}{[c/p-1]s + sF} , \text{ or}$$

$$\frac{d(1-\lambda)}{dp} = \frac{1}{c}\left(\frac{c}{p}\right)^2 \frac{(1-F)s}{c/p - 1 + F}\,\frac{1}{\sigma^2} .$$

Hence, along the $\pi = 0$ curve, for fixed p, provided all limits exist, we have

$$\lim_{\sigma\to\infty} \frac{d(1-\lambda)}{dp} = \frac{1}{c}\left(\frac{c}{p}\right)^2 \lim_{\sigma\to\infty}\frac{s}{\sigma^2} \cdot \lim_{\sigma\to\infty}\frac{1-F}{c/p - 1 + F}$$

$$= \frac{1}{c}\left(\frac{c}{p}\right)^2 \left[\lim_{\sigma\to\infty}\left[1 + \frac{u}{\sigma}\right]\right] \cdot 0$$

$$= \frac{1}{c}\left(\frac{c}{p}\right)^2 \left[\frac{p}{c}\right] \cdot 0$$

$$= 0 .$$

Q.E.D.

To prove the next proposition, we need the following lemma.

<u>Lemma 4:</u> The value of u, associated with any fixed $1 - \lambda$ (below 1) goes to ∞ as $\sigma \to \infty$, and, for fixed $1 - \lambda$, $\frac{u}{\sigma} \to -\lambda$ as $\sigma \to \infty$.

<u>Proof:</u> From (A8) we have that

$$(1 - \lambda) = \frac{I}{\sigma} + \frac{s}{\sigma^2} \quad \text{or}$$

$$-\lambda = \frac{I}{\sigma} + \frac{u}{\sigma} .$$

Differentiating the above, we obtain

$$\left.\frac{du}{d\sigma}\right|_{\lambda=\bar{\lambda}} = -\bar{\lambda}/1 - F(u) < 0$$ and is bounded away from zero for $1 - \lambda < 1$. Hence $u \to -\infty$ as $\sigma \to \infty$. Applying L'Hopital's rule to $\frac{u}{\sigma}$, one finds

$$\lim_{\sigma\to\infty} \frac{u}{\sigma} = \lim_{\sigma\to\infty} \frac{du}{d\sigma} = -\lambda \quad ,$$ Q.E.D.

The next lemma sees how along the $\pi = 0$ curve price behaves as σ increases, for fixed $1 - \lambda$.

<u>Lemma 5</u> Along the $\pi = 0$ curve, for fixed $1 - \lambda$ (below 1), the price, p, approaches c as σ increases.

<u>Proof:</u> From (A2), we have

$$p = \frac{cs}{\sigma I + s} \quad \text{which can be rewritten as}$$

$$p = \frac{c(\sigma^2 + u)}{(1 - \lambda)\ \sigma^2}$$

$$p = \frac{c}{1 - \lambda}\ (1 + \frac{u}{\sigma}) \quad .$$

For fixed $1 - \lambda$, as σ increases, $\frac{u}{\sigma} \to -\lambda$ from the previous lemma. Hence $p \to c$ as σ increases. Q.E.D.

Using the two preceeding lemmas, we can now prove:

<u>Proposition 7:</u> Along the $\pi = 0$ curve, for any fixed $1 - \lambda$ (below 1) the slope of the $\pi = 0$ curve becomes arbitrarily large as σ (or L/N) increases --

i.e.

$$\lim_{\sigma\to\infty} \left. \frac{d(1 - \lambda)}{dp} \right|_{\substack{\pi=0 \\ 1-\lambda=\overline{1-\lambda}}} \to \infty \quad \text{as} \quad \sigma \to \infty \; .$$

Proof: From (A11)

$$\frac{d(1-\lambda)}{dp} = \frac{1}{c}\,\frac{1}{\sigma^2}\,(1-F)\,\frac{(\sigma I + s)^2}{\sigma I + sF} \quad .$$

Since along the $\pi = 0$ curve we have $[\frac{c}{p} - 1]s = \sigma I$ and $(1-\lambda)\sigma^2 = \sigma I + s = cs/p$, we have that

$$\frac{d(1-\lambda)}{dp} = \frac{1}{c}\,\frac{1}{\sigma^2}\,(1-F)(1-\lambda)\sigma^2\,\frac{(c/p)s}{[c/p - 1]s + sF} \quad , \text{ or}$$

$$\frac{d(1-\lambda)}{dp} = \frac{(1-\lambda)}{c}\,(1-F)\,\frac{c}{p}\,\frac{1}{[c/p - 1]s + sF} \quad . \tag{A14}$$

From the preceding lemma we know that along the $\pi = 0$ curve, for fixed $1 - \lambda$, as σ increases, $u \to -\infty$ and $p \to c$. Hence,

$$\frac{d(1-\lambda)}{dp} \to \frac{(1-\lambda)}{c}\,\frac{1-F}{F} \to \infty \qquad \text{as } \sigma \text{ increases.} \qquad \text{Q.E.D.}$$

Proposition 8 As σ (or L/N) increases, the slope of the $\pi = 0$ curve falls at any fixed p.

Proof: It is necessary to show that $\frac{d}{d}(\frac{d\ 1-\lambda}{dp}) < 0$.

From (A14) we can write that

$$M \equiv \frac{d(1-\lambda)}{dp} = (1-\lambda)\ (-1)\ (\frac{c}{p})\ [\frac{1}{p - pF - c} + \frac{1}{c}] \ .$$

Consider $\frac{dM}{d\sigma} = \frac{d(1-\lambda)}{d\sigma}\ (-1)\ \frac{c}{p}[\frac{1}{p-pF-c} + \frac{1}{c}] + (1-\lambda)(-1)\frac{c}{p}\frac{(-1)(-p)f(u)}{[p-pF-c]^2}\cdot \frac{du}{d\sigma}$.

From Proposition 4, it follows that

$$\frac{d(1-\lambda)}{d\sigma} = \frac{c}{p}\quad \frac{1}{\sigma^2}\quad {}^{1/}{}_{1-c/p-F}(-f(u)) \tag{A15}$$

and

$$\frac{du}{d\sigma} = -(1-c/p)/_{(1-c/p-F)}\ .$$

Hence

$$\frac{dM}{d\sigma} = \frac{c}{p}\quad \frac{1}{\sigma^2}\quad \frac{1}{1-c/p-F}\ (-f)(-1)\ \frac{c}{p}\ [\frac{1}{p-pF-c} + \frac{1}{c}] +$$

$$+ (1-\lambda)(-1)\ \frac{c}{p}\ (-1)\ \frac{(-p)f}{(p-pF-c)^2}\ (-1)\ \frac{(1-c/p)}{(1-c/p-F)}\ ,$$

or

$$\frac{dM}{d\sigma} = c\ \frac{f}{(p-c-pF)^2}\ [(1-F)\cdot\frac{1}{\sigma^2} + (1-\lambda)\ \frac{(1-c/p)}{(1-c/p-F)}]\ .$$

So

$$\text{sgn}\ \frac{dM}{d\sigma} = \text{sgn}\ [\frac{(1-F)}{\sigma^2} + (1-\lambda)\ \frac{(1-c/p)}{(1-c/p-F)}] \equiv \text{sgn A}$$

To prove that the expression in brackets above is negative, we will proceed in two steps. First, we will show that for any p, the expression equals zero as $\sigma \to \sigma_{min}$ where σ_{min} is the smallest σ for which the $\pi=0$ curve exists for that ρ. Then we will show that the expression in brackets is monotone increasing in σ.

First, let $\sigma \to \sigma_{min}$ = $1-c/p-F \to 0$ so $1-c/p \to F$. Then,

$$\lim_{\sigma \to \sigma_{min}} A \equiv \lim_{\sigma \to \sigma_{min}} [(\frac{1-F}{\sigma^2}) + (1-\lambda)(1-c/p/_{1-c/p-F}] \to$$

$$c/p \cdot \frac{1}{\sigma^2} + (1-c/p) \cdot \ell \qquad \text{where}$$

$$\ell = \lim_{\sigma \to \sigma_{min}} (1-\lambda)/_{1-c/p-F} \, .$$

Using L'Hopital's rule, we find

$$\ell = \frac{\frac{d(1-\lambda)}{d\sigma}}{\frac{d(1-c/p-F)}{d\sigma}} = \frac{\frac{c}{p}\,\frac{1}{\sigma^2}\,\frac{1}{1-c/p-F}\,(-1)f}{(-1)f\,\frac{du}{d\sigma}}$$

or

$$\ell = \frac{\frac{c}{p}\,\frac{1}{\sigma^2}\,\frac{1}{1-c/p-F}\,(-1)f}{(-1)f\,(-1)\,(1-c/p)/(1-c/p-F)} \, ,$$

or

$$\ell = (-1)\,\frac{c}{p}\,\frac{1}{\sigma^2}\,\frac{1}{1-c/p} \, .$$

Plugging in for ℓ in the expression for A we find

$$\lim_{\sigma \to \sigma_{min}} A = [\frac{c}{p}\frac{1}{\sigma^2} - \frac{c}{p}\frac{1}{\sigma^2}] = 0 .$$

We now wish to show that A decreases as σ increases above σ_{min}. This is equivalent to showing that

$$\frac{d[(1-\lambda)\sigma^2]}{d\sigma} > \frac{d}{d\sigma}[(-1)\frac{(1-F)(1-c/p-F)}{(1-c/p)}] . \quad (A16)$$

Now,

$$\frac{d(1-\lambda)}{d\sigma} = \sigma^2 . \frac{d(1-\lambda)}{d\sigma} + 2(1-\lambda) > \sigma^2 \frac{d(1-\lambda)}{d\sigma} .$$

From (A15) we have that

$$\sigma^2 \frac{d(1-\lambda)}{d\sigma} = \frac{c}{p}(-1)\frac{f}{1-c/p-F} . \quad (A17)$$

Now consider the RHS of (A16):

$$\frac{d}{d\sigma}(-1)\frac{[[1-F]^2 - c/p(1-F)]}{1-c/p} = \frac{-1}{1-c/p}[2(1-F)(-f) + (c/p)f]\frac{du}{d\sigma} ,$$

$$= \frac{-1}{1 - c/p} \, [2(1 - F)(-f) + c/p \cdot f] \, (-1) \, \frac{1 - c/p}{1 - c/p - F}$$

$$= \frac{(-f)}{1 - c/p - F} \, [2 - 2F - c/p] \qquad \text{(A18)} \; .$$

To see if (A17) > (A18), we ask whether

$$\frac{c}{p} > 2 - 2F - c/p \quad \text{, or}$$

$$\frac{c}{p} > 1 - F \quad \text{or}$$

$$1 - \frac{c}{p} - F < 0$$, which is always true, as discussed earlier.

This proves that the inequality in (A16) is always true for $\sigma > \sigma_{min}$, and hence the proof of the theorem follows. Q.E.D.

From (11), we have

$$\frac{d\, 1 - \lambda}{dp} = \frac{1}{c} \frac{(\sigma I + s)^2}{(\sigma I + sF)} \frac{1}{\sigma^2} (1 - F) \quad ,$$

or using (A8) ,

$$\frac{d\, 1 - \lambda}{dp} = \frac{1}{c} (1 - \lambda)^2 \; \frac{1}{(1 - \lambda) - \frac{s}{\sigma^2}(1 - F)} \; (1 - F) \quad .$$

Therefore, as L/N increases, the slope at equilibrium along the $\pi = 0$ curve equals

$$\frac{d\,\overline{1-\lambda}}{dp} \approx \frac{1}{c}\,\frac{1 - F(u)}{1 - \frac{s}{\sigma^2}(1 - F(u))} \; .$$

In Chapter 2 we prove that as L/N increases, the equilibrium price p approaches c and the probability of satisfaction, $1 - \lambda$, approaches 1. From Theorem 3 of Chapter 2, we know that in equilibrium $\frac{s}{\sigma^2} \to 1$ as L/N increases, so we have that at the equilibrium point, as L/N increases,

$$\frac{d\,\overline{1-\lambda}}{dp} \approx \frac{1}{c}\,\frac{1 - F(u)}{F(u)} \; .$$

In Chapter 2 we will need to determine the smallest value for σ^2 (i.e. L/N) such that it is possible for equilibrium to involve a price below $\overline{p}$, and a probability of satisfaction above $\overline{1 - \lambda}$. To solve for the desired σ^2, we need to solve the following two simultaneous equations in σ and s:

$$\pi\,(1 - \lambda, \overline{p}) = 0 \quad \text{and}$$

$$1 - \lambda = \overline{1 - \lambda} \quad , \text{ or}$$

from (A1) and (A8) ,

$$\frac{(\sigma I + s)}{\sigma^2} = \overline{1 - \lambda} \quad \text{and} \tag{A19}$$

$(\sigma I + s)\ \overline{p} \cdot L - NCs = 0$ or equivalently

$$\frac{cs}{\sigma I + s} = \overline{p} \ . \tag{A20}$$

Solving the system (A19) and (A20) is equivalent to solving

$$(\overline{1 - \lambda})\ \frac{(\overline{p})}{c} = \frac{s}{\sigma^2} \quad \text{and} \tag{A21}$$

$$\frac{\sigma I + s}{\sigma^2} = \overline{1 - \lambda} \ . \tag{A22}$$

In general, to solve (A21) and (A22) for σ and s, it is necessary to use a computer algorithm. However, in one case, the solution works out very nicely. Suppose $(\overline{1 - \lambda})\ (\frac{\overline{p}}{c}) = 1$, then $\frac{s}{\sigma^2} = 1$. Since $u \equiv \frac{s - \sigma^2}{\sigma}$ we see that u will equal 0 and hence $I(u)$ is independent of s and σ and equals $I(0) = -\sqrt{\frac{2}{\pi}}$. For this case, we can then solve (A22) to find

$$\frac{I(0)}{\sigma} + 1 = \overline{1 - \lambda} \quad \text{or}$$

$$\sigma^2 = \frac{[I(0)]^2}{\overline{\lambda}^2} \ . \tag{A23}$$

In Chapter 2 we use (A23) to solve for the σ^2 that satisfies the system of equations (A19) and (A20).

CHAPTER 3

Models With Price Taking Behavior: Fair Trade and Monopoly

3.0 Introduction

In the previous chapter, we saw that firms in competition will behave as "utility-level takers." In this short chapter, we construct models of market behavior in which we assume that either the information flows and/or the institutional structure are such that a firm will behave as either a price taker or as a monopolist. First we assume that the firm purchases the good from a wholesaler, who dictates the retail price that the firm must charge. For expositional purposes we refer to this model as a fair trade model. If we assume that wholesalers compete with each other, then we show how competition among the wholesalers will establish an equilibrium, whose properties we compare to those of the competitive equilibrium derived in Chapter 2. We then briefly examine the behavior of a monopolist. The response of a price taking firm to an unexpected increase in uncertainty is compared to that of a monopolist. The effect of these responses on the ability of an economy to function smoothly in changing times is discussed.

3.1 A Model With Price Taking Behavior

The model is very similar to that of the previous chapter. There are L identical customers and N identical firms. During a market period, each consumer visits one firm to try to satisfy his demand for the product. Consumers will randomly frequent any firm that they feel is offering the highest level of expected utility. At the beginning of each market period, firms choose a wholesaler and decide how much to purchase from that wholesaler at price c. Unlike the model of the previous chapter, in this model firms are told by the wholesaler what retail price to charge for the good. Just as in the previous chapter, the good is perishable, no deliveries of the good occur during a market period, and a firm's price does not vary during a market period. As before, natural features of market operation are that some customers are unable to purchase the product and some firms are unable to sell all their goods.

The feature that distinguishes this model from the one of the previous chapter deals with the information that consumers have available. In this model, consumers are able to distinguish firms only by the identity of their wholesaler. We imagine that consumers can gain information on the identity of the wholesaler who has supplied the firm which is offering the best deal in the market. In other words, some consumer group determines who the best firm is, but only dispenses information on the identify of the wholesaler of this firm. In contrast to the model of the previous chapter, consumers do not know what level of utility the best firm is offering.

3.2 Behavior of the Firm

For the moment, let us assume that all firms buy from the same wholesaler and that the retail price is fixed at p. Firms will receive randomly $\frac{1}{N}$'th of the L customers each period, and will want to choose their customers capacity, s, so that profits are maximized at that price. From equation (3) of Chapter 2, we know that profits are some function $\pi(s,p)$ of price, p, and customer capacity s. For a given price a firm will choose s so that the following first order condition is satisfied:

$$\pi_s(s,p) = 0 \tag{1}$$

where a subscript denotes partial differentiation.[1] In the appendix to this chapter, it is shown that (1) can be approximated as

$$1 - F(n_c) = c/p \tag{2}$$

where F = cumulative normal distribution,

$$n_c = \frac{s - \sigma^2}{\sigma},$$

and $\sigma^2 = L/N$.

The customer capacity, s, and the probability of satisfaction, $1 - \lambda$, are related by

$$1 - \lambda = \frac{\sigma I(u) + s}{\sigma^2} \tag{3}$$

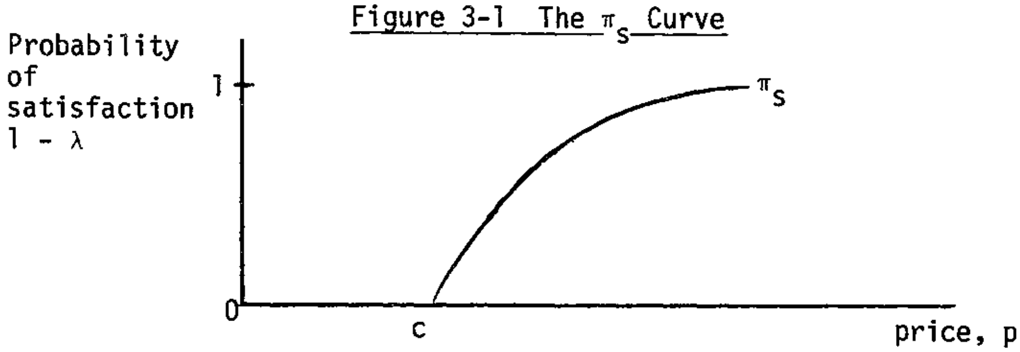
where (3) was derived previously in Appendix A, and where

$$I(u) = -f(u) - uF(u),$$

and $f(u)$ = normal density function.

The relations (2) and (3) establish the locus of points in $(1 - \lambda, p)$ space that satisfy the first order conditions. The shape of this locus of points, which we henceforth call the π_s curve, is depicted below.

Figure 3-1 The π_s Curve



[1] In the appendix to this chapter, it is shown that the second order conditions are satisfied so that (1) defines an interior maximum.

The geometrical properties of the π_s curve are very similar to those of the zero profit (π = 0) curve of the previous chapter, and are derived in the appendix to this chapter. The π_s curve is concave (i.e., $\frac{d^2(1-\lambda)}{dp^2} < 0$) and rises from 0 to 1 as price increases. As the customer per firm ratio, L/N, increases, the slope for any fixed price (greater than c) falls to zero, while the slope for any fixed probability of satisfaction (below 1) becomes arbitrarily large. The price on the horizontal axis at which the π_s curve begins to rise approaches c as L/N increases. From Figure 2-2 of Chapter 2, we see that the π_s curve lies entirely below the $\pi = 0$ curve used in the previous chapter to define the competitive equilibrium.

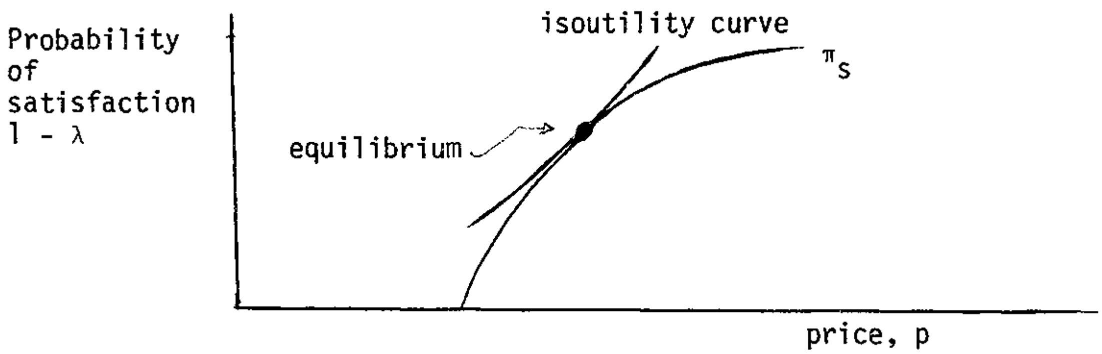
Given the price, p, that the wholesaler dictates, the π_s curve tells us what probability of satisfaction a firm will offer. We now wish to investigate how the wholesalers will determine the mark-up that they require firms to charge, if wholesalers compete among themselves.

3.3 The Fair Trade Equilibrium

Suppose that there are several wholesalers of the good. Each wholesaler charges c for the good and dictates the price, or mark-up, that firms who buy from him must charge. Consumers will only want to frequent firms whose wholesalers allow them to charge a mark-up so that the firm, when operating on the π_s curve, is able to offer the highest possible utility level to consumers. Competition among wholesalers replaces competition among firms as the force that establishes market equilibrium. In this model, firms adopt the passive role of choosing the best wholesaler, and operating on the π_s curve at the price that the wholesaler establishes.

Wholesalers will compete on the mark-up until the utility level, associated with firms who operate on the π_s curve at that mark-up, could not be improved if a different mark-up were charged. Under the incentives just discussed, the fair trade equilibrium occurs at the tangency between the π_s curve and the highest possible isoutility curve, as depicted below.

Figure 3-2 The Fair Trade Equilibrium



As in the previous chapter, it is possible to construct a simple story of how the dynamics of the market lead us to the equilibrium point. The story is much the same as before, although the individual firms now play a relatively passive role in the determination of equilibrium. Because of the similarity between the two stories, there is no need to repeat the reasoning underlying the basic features of the story.

Firms buy from the same wholesaler for a one week period. During this week, neither the wholesaler, nor the firm alters their operating policies. Firms operate on the π_s curve at the price that their wholesaler dictates. Initially, customers feel that all firms are identical and shop randomly among them for a one week period, visiting one firm per day. At the end of the week, it becomes known (through some consumer groups) who the wholesaler is that supplies the firm who offered the best deal (i.e., the highest level of expected utility) for this past week. Wholesalers realize that to prevent firms from switching to this best wholesaler for their supplies for next week, wholesalers must alter their mark-up policy, so that firms, whom they supply with goods, will at least offer this highest level of expected utility. Wholesalers spend money on some sort of promotional scheme to reestablish their competitiveness in the minds of consumers. Wholesalers also publish a list of firms that next week will buy from them. As in the previous chapter, we assume that consumers will never return to a firm whose wholesaler does not live up to his promises - so that false advertising does not occur.[1]

Once the competitiveness of each wholesaler is reestablished in the minds of consumers (remember consumers can only distinguish firms by the identity of their wholesaler), consumers once again shop randomly among the firms next week. Next week, firms once again take the wholesaler's price as given and choose their optimal policies according to the π_s curve. In this scenario, wholesalers have a posivitive incentive to offer that mark-up that causes firms to adopt policies that provide the highest utility level to consumers. These incentives will cause the fair trade equilibrium depicted in Figure 3-2 to be established.

3.4 Properties of the Fair Trade Equilibrium

The qualitative differences between the fair trade equilibrium and the corresponding deterministic equilibrium are identical to the differences between the competitive equilibrium of the previous chapter and the deterministic equilibrium. Since these differences were discussed in detail in the previous chapter, they will no be repeated here. Because of the similarity in the geometrical properties of the π_s and $\pi = 0$ curve, the behavior of the fair trade equilibrium as the customer per firm ratio increases is

[1]As in the previous chapter, firms are assumed to desire to stay in business for more than one week.

similar to that of the competitive equilibrium. In particular, using proofs similar to those in the previous chapter, the following theorem holds.

Theorem 1: As the customer per firm ratio increases:

1. the fair trade equilibrium price approaches c,
2. the fair trade equilibrium probability of satisfaction approaches 1, and
3. in the fair trade equilibrium, the percent discrepancy between supply and demand approaches zero.[1]

There are however significant differences between the fair trade equilibrium and the competitive equilibrium of the previous chapter. The π_S curve lies entirely below the $\pi = 0$ curve used to define the competitive equilibrium. Hence compared to the competitive equilibrium, the fair trade equilibrium will involve a higher (i.e. positive) level of profits[2] to firms and a lower of utility to consumers. The fair trade equilibrium will usually involve a higher price and lower probability of satisfaction than the competitive equilibrium.[3]

Since the π_S curve lies below the $\pi = 0$ curve, it follows that, in general, the customer per firm ratio in the fair trade case will have to be larger than it was in the competitive case, if the deterministic analysis (i.e. price equals c, probability of shortage equals 0) is to meet desired tolerance limits. For example, we can ask the same question as in the previous chapter of how large does the customer per firm ratio have to

[1]The proof of this statement follows immediately from part 2 of Theorem 1 and (3).

[2]Since profits are positive, we would expect entry of additional firms to occur in the long run. The situation above with positive profits and a fixed number of firms, corresponds to a short run equilibrium.

[3]"Usually" because we are assuming that when a consumer has a better menu of $(1 - \lambda, p)$ choices available, he will choose to be made better off in each dimension. Note the analogy to a "normal" income effect in a two good model.

be before the deterministic analysis can even hope to meet desired tolerance levels.[1] If we are willing to tolerate only 1% errors in price and probability of shortage, then the minimum customer per firm ratio that is required before the deterministic analysis can even hope to be approximately correct is 62,000.[2] If we relax the tolerance limits to 2% errors in the price and probability of satisfaction, then this minimum customer per firm ratio is 10,000, still a very large number. The corresponding minimum customer per firm ratios for the competitive equilibrium of the previous chapter were 6500 for 1% error tolerances and 1600 for 2% error tolerances. Even more so than in the previous chapter, for most purposes, the customer per firm ratio must be unrealistically large before the deterministic analysis, which ignores the uncertainty in the market, can be used to predict the fair trade equilibrium price and probability of satisfaction. Moreover, as in Chapter 2, even in the case of a very large customer per firm ratio, the deterministic analysis will be unable to detect or predict the consequences of having supply being unequal to demand by arbitrarily large amounts in the fair trade equilibrium.

Another difference between the competitive and fair trade equilibrium is that unlike the competitive equilibrium, the fair trade equilibrium is not Pareto Optimal in the sense that both firms and consumers can be made better off. All isoprofit curves that cross the π_s curve do so with a vertical slope. Therefore, invoking the upper bound (i.e. Lipschitz condition) on the slope of the isoutility curves, it is clear that the π_s curve can never cross the contract curve defined by the isoutility and isoprofit curves. Hence, the fair trade

[1] See Chapter 2, Section 2.6.5 for a discussion of the reasoning underlying this question.

[2] This figure is computed from Table B-1 and relation (B9) which appear in the appendix to this chapter.

equilibrium will in general not lie on the contract curve, and therefore, there will usually exist Pareto superior points to the fair trade equilibrium. (The only exception to this statement occurs for monopoly pricing, which we discuss in the next section.)

A final difference between the fair trade and competitive equilibrium deals with the absolute discrepancy between supply and demand, as the customers per firm ratio increases. In the competitive equilibrium of the previous chapter, it is possible for supply to either exceed or fall short of demand by arbitrarily large amounts as the customer per firm ratio increases. For the fair trade equilibrium, supply will always fall short of demand by arbitrarily large amounts as the customer per firm ratio increases. To see this last point, recall that the fair trade equilibrium price approaches c as the customers per firm ratio rises. From (2), this last statement implies that n_c approaches $-\infty$, which, in turn, implies that in the fair trade equilibrium supply falls short of demand by arbitrarily large amounts as the customer per firm ratio increases.

3.5 Monopoly

Since the case where firms take price as fixed was examined in Section 3.2, it is relatively straightforward to extend the analysis to the case of a monopolist.[1] So, suppose a monopolist who owns all N firms wants to choose the same price and stocking policy for each firm so as to maximize expected profits. As usual, firms are not allowed to ship the goods amongst themselves during a market period.

The monopolist wants to choose customer capacity, s, and a price, p, to maximize profits, $\pi(s,p)$. In addition to the first order condition (1),

[1]The original study of monopoly under uncertainty was done by E. Mills, "Uncertainty and Price Theory", Quarterly Journal of Economics, 1959.

the monopolist will set $\pi_p(s,p) = 0$, where the subscript denotes partial differentiation. (We assume the second order conditions for a maximum are met.)

The monopolist operates at that point on the π_s curve of Figure 3-1 that maximizes profits. Notice that this point must lie on the contract curve between the isoutility and isoprofit curves, since profits are maximized at this point. Unlike either the competitive or fair trade equilibrium, the monopoly equilibrium does not take into account consumer preferences for the probability of satisfaction. Hence, the level of utility in the fair trade and competitive equilibria will exceed that in the monopoly equilibrium. The only exception to this last statement occurs in the fair trade equilibrium when the isoutility curves happen to be tangent to the π_s curve precisely at the monopoly point. In that case, the fair trade and monopoly equilibria are identical.

The random demand that the monopolist observes at each firm equals $x(p) \cdot i$ where i is the random number of customers who frequent the store in any period. Since the randomness is multiplicative, we know from Karlin and Carr[1] that the price the monopolist charges will exceed the price he would charge in the corresponding riskless environment. Of course, in the riskless environment, the monopolist's price always exceeds the constant marginal cost c. From the results of this and the previous chapter, for sufficiently large customer per firm ratios, we can be sure that the monopolist's price will exceed the price in both the competitive and the fair trade equilibria. Since the π_s curve rises as price increases, it also follows that for a sufficiently large customer per firm ratio, the

[1] S. Karlin and C. Carr, "Prices and Optimal Inventory Policy", in K. Arrow, S. Karlin, and H. Scarf, Studies in Applied Probability and Management Science, Stanford University Press, 1962.

probability of satisfaction will be higher under monopoly than it is in the fair trade equilibrium, and can even be higher than it is in the competitive equilibrium. These last results are interesting because monopoly is usually associated with a restriction of output, so that intuitively we might have expected that the probability of satisfaction would always be lower in the monopoly equilibrium than in either the fair trade or competitive equilibria.

3.6 Unexpected Increases in Uncertainty and A Smoothly Functioning Economy

So far, we have been examining the situation where a firm cannot exactly predict its demand, but instead knows the distribution of the random demand, which in turn influences its operating policies. One question that arises is what happens if the environment suddenly becomes more uncertain. If prices are endogenous, then supposedly we will immediately move to a new price-probability of shortage equilibrium. On the other hand, in the very short run, it may be reasonable to suppose that prices remain fixed at their current levels, and that only quantity adjustments can occur. Customers may dislike price changes or it may be costly for the firm to change prices, so that firms may resist changing price until they are convinced that the shift in the stochastic structure of demand is permanent. In this very short run, then, firms can alter only their stocking policies in response to the unexpected increase in uncertainty.

In this very short run, an economic system, if it is to be a viable system, must respond in such a way that the economy can continue to function smoothly. For example, the short run response of firms to this increase in demand uncertainty must not be to close up shop and cease production until they can determine whether the unexpected shift in the stochastic structure of demand is permanent. Other firms who need to rely on the firms who have stopped production for input supplies would be forced to curtail production of their goods, and consumers would be forced to do without the goods. Any economic system must have enough flexibility so that, in these very short run periods of quantity adjustments, chaotic increases in the severity of shortages do not occur.

Consider the following situation. Firms are behaving as either price takers or monopolists and the market is in one of the equilibria, described previously in this chapter. Suddenly, there is an unexpected increase in the uncertainty of demand, so that the expected shortage rises by some amount, y. In the very short run prices are fixed, and firms can only adjust the amount of the goods that they stock. The question we want to answer is whether in this very short run period the firm will try to offset the increase, y, in the expected shortage by stocking more goods, or whether the firm will do just the oppposite, and exacerbate the increase in the expected shortage by cutting back on the amount of goods that it holds.

The first response of mitigating the increased shortage is desirable from the point of view of preserving the smooth functioning of the economy. The second response of worsening the increased shortage would appear to be undesirable since it could result in severe shortages that could force drastic curtailments of production elsewhere in the economy. The next theorem answers the question of when we can expect firms to adopt the stabilizing response which mitigates the increased shortage, and when we can expect firms to adopt the destabilizing response which worsens the increased shortage.

Theorem 2: For fixed price, and assuming a symetric random distribution of demand, if $p > 2c$ ($p < 2c$), then firms increase (decrease) their stock in response to an unexpected increase in uncertainty.

Proof: Since demand is uncertain and price is fixed we can write that the demand x equals

$$x = \bar{x} + e,$$

where $\bar{x}$ = expected demand at the fixed price p,

and e = error term with expected value of 0, and symetric density $f_1(e)$, and cumulative density $F_1(e)$.

We now suppose that the error term remains symetric but becomes riskier in the Rothschild-Stiglitz sense[1] and is described by the cumulative distribution function $F_2(e)$. From symetry and from the definition of increasing risk, it immediately follows that

$$F_1(0) = F_2(0), \qquad (4)^{2}$$

$$F_1(e) \geq F_2(e) \quad \text{for } e \geq 0, \qquad (5)$$

$$\text{and } F_1(e) \leq F_2(e) \quad \text{for } e \leq 0. \qquad (6)$$

At fixed price, p, the first order condition that determines the optimal stock, S, is derived in the appendix to this chapter and is given by

[1] M. Rothschild and J. Stiglitz, "Increasing Risk: I, A Definition," Journal of Economic Theory, 1970.

[2] This is the only place where the symetry assumption on the distribution of the random term, e, is used. As is clear from (4), this symetry assumption could be weakened to an assumption requiring the mean and median of the distribution of demand to be equal, and the theorem and its proof would be unchanged.

$$F_i(S_i - \bar{x}) = 1 - c/p, \qquad (7)$$

for i = 1, or i = 2.

It follows from (4) through (7) that if $1 - c/p < 1/2$. then $S_2 < S_1$, while if $1 - c/p > 1/2$, $S_2 > S_1$. Q.E.D.[1]

We have already mentioned that the monopolist's price exceeds the riskless monopoly price. If we assume that the price elasticity of demand is between -1 and -2, then we find that for the monopoly situation, $p > 2c$. For this case, monopoly always provides a stabilizing response in the very short run, when prices are fixed, in reaction to unexpected increases in uncertainty. In contrast, in the fair trade equilibrium, for a sufficiently large customer per firm ratio, price taking firms will always provide a destabilizing response in reaction to unexpected increases in uncertainty in the very short run when prices are fixed. Therefore, compared to the fair trade equilibrium, monopoly usually will involve welfare losses to customers in a static situation. However, in the case where the price elasticity of demand is between -1 and -2, monopoly does have the virtue that in the very short run with fixed prices, its responses to unexpected increases in uncertainty will definitely be to stabilize the functioning of the economy.

3.7 Summary

This chapter has analyzed models of market clearing under uncertainty in which the firm behaves either as a price taker or as a monopolist. For a particular flow of information, it was shown how the fair trade equilibrium would be established.

[1] It is obvious from (7) that the condition $p \gtrless 2c$ also determines whether supply, s, exceeds expected demand, $\bar{x}$, for the models of this chapter.

The qualitative differences between the deterministic equilibrium and either the fair trade or monopoly equilibria are similar to those between the deterministic equilibrium and the competitive equilibrium of the previous chapter. However, there are significant differences between the competitive, fair trade and monopoly equilibria. The competitive equilibria will always involve a higher level of expected utility than either of the other two equilibria. The fair trade equilibrium is usually not Pareto-Optimal from the consumer and firm point of view. For a sufficiently large customer per firm ratio, the monopoly equilibrium will definitely involve a higher price than either of the other two equilibria. The probability of satisfaction in the monopoly equilibrium will exceed that in the fair trade equilibrium and may exceed that in the competitive equilibrium, for a sufficiently large customer per firm ratio. It was found that in a short run environment of fixed prices, a monopolist, and not price-taking firms in the fair trade equilibrium, will be more likely to adopt "stabilizing" operating policies that will partially offset any increase in expected shortages caused by unforeseen increases in demand uncertainty.

APPENDIX B

Mathematical Appendix to Chapter 3

In this section, we examine firm response when the price (> c) is taken as given. From Chapter 3, we have that

$$\pi(s) = [p \sum_{0}^{s} i\ pr(i) + ps \sum_{s+1}^{\infty} pr(i) - cs]x(p),$$

where pr(i) = binomial probability of obtaining i customers, and all other variables were previously defined in Chapters 2 and 3. Maximizing with respect to s, we obtain the first order condition

$$\sum_{s+1}^{L} pr(i) = c/p.$$

Henceforth, we assume the above first order condition holds with equality for some positive s.[1] Since it is obvious that the second order conditions are met, the above first order condition defines an interior maximum.[2]

Let $\bar{s}$ = L/N = mean of the binomial process, and $\sigma = \sqrt{L/N}$ = approximate variance of the binomial process. As before, f and F represent the density and cumulative density for the normal distribution. Define n_c by the equation $1 - F(n_c) = c/p$. Then, we can rewrite the above first order condition as

$$s \approx \bar{s} + \sigma \cdot n_c. \tag{B1}$$

---

[1]If price is sufficiently low, the firm will not produce, while if price is sufficiently high, the firm will produce enough to satisfy the entire market by itself. For the remainder of this appendix, we will only be concerned with the price range over which price is sufficiently high to justify positive production, yet not so high to justify one firm from producing enough to satisfy the entire market. To avoid tedious repetition, we will not continue to mention this fact.

[2]I.e., $\Delta^2 (s) = -p\,pr(s) < 0$, where Δ^2 indicates a second difference.

Since n_c is a function of p, (B1) expresses s as a function of p. We now want to examine the relation between the probability of satisfaction $1 - \lambda$ and the customer capacity s. The expected shortage, M, at a firm with customer capacity s equals

$$M = \sum_{s+1}^{L} (i - s)\, pr(i),$$

or

$$M = \sum_{s+1}^{L} (i - s)\, pr(i) + (\bar{s} - s) \sum_{s+1}^{L} pr(i),$$

or since $\frac{i - \bar{s}}{\sigma}$ has an approximately N(0,1) distribution, we have

$$M = \sigma N^{Er}\left(\frac{s - \bar{s}}{\sigma}\right) + (\bar{s} - s)\left[1 - F\left(\frac{s - \bar{s}}{\sigma}\right)\right],$$

where $N^{Er}(x) \equiv \int_x^\infty tf(t)dt = f(t)$. Using (B1), we have

$$M = \sigma\left[N^{Er}(n_c) - n_c[1 - F(n_c)]\right] ,$$

or

$$M = \sigma\left[f(n_c) - n_c[1 - F(n_c)]\right] . \qquad \text{(B2)}$$

Since n_c is a function of price (since $1 - F(n_c) = c/p$), (B2) expressed the expected shortage as a function of price. We want to establish how the expected shortage behaves as a function of price. First, we prove a lemma.

Lemma 1: Consider the n_c that satisfies $F(n_c) = 1 - c/p$. The quantity dn_c/dp is positive and equals $(1 - F(n_c))^2/(c \cdot f(n_c))$.

Proof: By definition

$$\int_{n_c}^\infty f(u)du = c/p,$$

which upon differentiation yields,

$$\frac{dn_c}{dp} = \frac{c}{p^2} \cdot \frac{1}{f_n(n_c)} = \frac{1}{c}\left(\frac{c}{p}\right)^2 \cdot \frac{1}{f_n(n_c)},$$

or since $F(n_c) = 1 - \frac{c}{p}$, we have that

$$\frac{dn_c}{dp} = [1 - F(n_c)]^2 \cdot \frac{1}{c \cdot f(n_c)} > 0. \qquad \text{Q.E.D.}$$

The fact that dn_c/dp is positive is equivalent to the statement that the optimal number of customers to service is monotone increasing in p. We can now prove:

Proposition 1: The expected shortage, M, is monotone decreasing in price, p.

Proof: Differentiate (B2) to find that

$$\frac{dM}{dn_c} = \sigma[f'(n_c) - 1 + F(n_c) + n_c \cdot F'(n_c)],$$

or , since $f'(n_c) = -n_c f(n_c)$ and $F'(n_c) = f(n_c)$,

$$\frac{dM}{dn_c} = \sigma[-n_c f(n_c) - 1 + F(n_c) + n_c f(n_c)],$$

or

$$\frac{dM}{dn_c} = \sigma[1 - F(n_c)](-1),$$

or

$$\frac{dM}{dn_c} < 0 \qquad \text{(B3)}$$

Now

$$\frac{dM}{dp} = \frac{dM}{dn_c}\frac{dn_c}{dp}.$$

From Lemma 1, we know that dnc/dp is positive, so that it follows from (B3) that dM/cp is negative. Furthermore, we see that

$$\frac{dM}{dp} = \frac{dM}{dn_c}\frac{dn_c}{dp} = \sigma[1 - F(n_c)]^3 \frac{1}{c \cdot f(n_c)} (-1) < 0. \quad \text{Q.E.D.} \tag{B4}$$

Proposition 1 implies that the expected number of customers who are disappointed by the firm's optimal servicing policy declines as p increases since each firm services more customers as p increases. A much more important property of the function M(p) has to do with its convexity.

<u>Proposition 2</u>: The function M(p) is convex in p.

<u>Proof</u>: First compute $d[dM/dp]/dn_c$ from (B4).

$$\frac{d(\frac{dM}{dp})}{dn_c} = (-1)\frac{1}{c}\ \sigma\left[\frac{3[1 - F(n_c)]^2(-1)f\ (n_c)}{f(n_c)} - \frac{[1 - F(n_c)]^3 f'(n_c)}{f(n_c)^2}\right],$$

or $$\frac{d(\frac{dM}{dp})}{dn_c} = (-1)\frac{1}{c}\ \sigma\ [3(1 - F(n_c))^2(-1) + \frac{(1 - F(n_c))n_c}{f(n_c)}],$$

or $$\frac{d(\frac{dM}{dp})}{dn_c} = \frac{1}{c}\ \sigma[1 - F(n_c)]^2\ [3 - \frac{(1 - F(n_c)) \cdot n_c}{f(n_c)}].$$

Now $$\frac{d^2M}{dp^2} = [d(dM/dp)/dn_c]\frac{dn_c}{dp}\ .$$

Since from Lemma 1 we know that sgn dn_c/dp is positive, it follows that

$$\text{sgn}\frac{d^2M}{dp^2} = \text{sgn}[3 - \frac{(1 - F(n_c))n_c}{f(n_c)}]. \tag{B5}$$

I will now proceed to show that the expression in (B5) is positive for all n_c. There are two cases to consider, corresponding to $n_c \gtrless 0$.

If $n_c < 0$, it is obvious that the expression in (B5) is positive. If $n_c > 0$, then from Birnbaum's Inequality,[1] it follows that

$$\frac{[1 - F(n_c)]n_c}{f(n_c)} < 1,$$

which implies that the expression in (B5) is positive. Therefore

$$\frac{d^2M}{dp^2} > 0$$

for all n_c, or equivalently for all p for which the curve exists (i.e., for all p which justify positive production). Q.E.D.

When all firms behave alike, the total expected shortage is $N \cdot M$. The fraction of customers who are dissatisfied is then $\frac{N \cdot M}{L}$, so from (B2) the probability of satisfaction is

$$1 - \lambda = 1 - \frac{1}{\sigma}\left[f(n_c) - [1 - F(n_c)]n_c\right] \qquad \text{(B6)}.$$

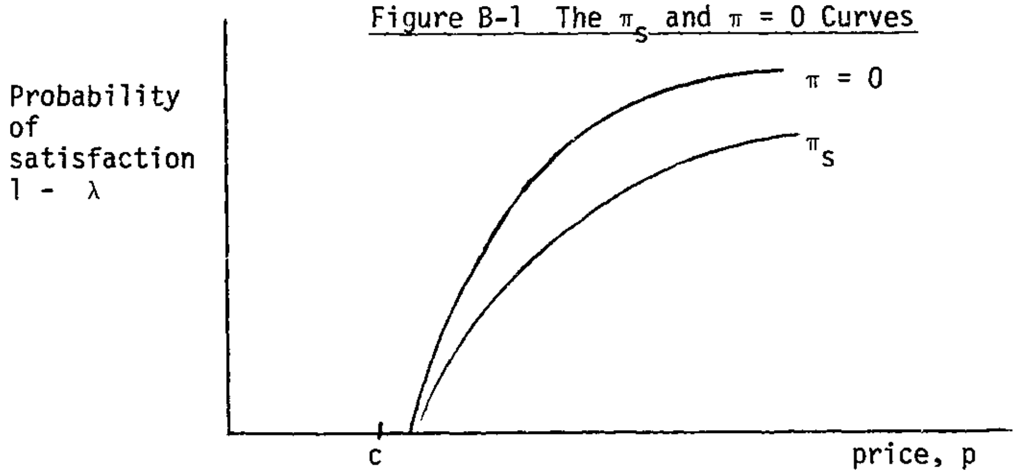
The above relation expresses $1 - \lambda$ as a function of p, since n_c is a function of p. It immediately follows from Proposition 2 that $1 - \lambda$ is concave in p over the relevant point range. Let us call this curve in $(1 - \lambda, p)$ space the π_s curve.

Notice that along the π_s curve at any price p, $1 - F(n_c) = c/p$, while in Appendix A we saw that along the $\pi = 0$ curve, $1 - F(u) - c/p < 0$. It therefore follows that at price p, $u > n_c$. Since $s = \bar{s} + u\sigma$ or $s = \bar{s} + n_c\sigma$, it follows that at any p for which the two curves exist, the associated value of s is larger along the $\pi = 0$ curve than along the π_s curve. Since

[1] M. Kendall and G. Stuart, The Advanced Theory of Statistics, Vol. 1, p. 147. - Birnbaum's Inequality states that $\frac{1 - F(x)}{f(x)} < \frac{4}{3x + \sqrt{x^2 + 8}}$, $x \geq 0$.

the probability of shortage is monotonic in s, we obtain the result that the π = o curve lies above the π_s curve over the relevant price range. The diagram below illustrates these points.

Figure B-1 The π_s and π = 0 Curves



Let us now investigate how the π_s curve is affected as σ (or L/N) increases. Rewrite (B6) as

$$1 - \lambda = 1 - \frac{1}{\sigma} H(n_c) \qquad \text{(B7)},$$

where $$H(n_c) = f(n_c) - n_c[1 - F(n_c)].$$

Since n_c does not depend on σ, we immediately obtain:

Proposition 3: For fixed p, the π_s curve shifts up to $1 - \lambda = 1$ as σ (or L/N) increases.

Proposition 4: For fixed p, the slope $\frac{d(1-\lambda)}{dp}$ along the π_s curve goes to zero as σ (or L/N) increases.

We now want to investigate the slope $\frac{d(1-\lambda)}{dp}$ along the π_s curve when $1 - \lambda$ is held fixed. From Lemma 4 of Appendix A, we know that $n_c \to -\infty$ as σ increases for fixed $1 - \lambda$ (below 1). Since $F(n_c) = 1 - c/p$ along the π_s curve, we see that:

Lemma 2: The price, associated with a fixed $1 - \lambda$, along the π_s curve approaches c as σ (or L/N) increases.

We can now prove:

Proposition 5: For fixed $\overline{1 - \lambda}(< 1)$, the slope, $\frac{d(1 - \lambda)}{dp}$, along the π_s curve goes to infinity as σ (or L/N) increases.

Proof: From above, we know that $n_c \to -\infty$ as σ increases. Since $f(n_c) = F(n_c) \to 0$ as $n_c \to -\infty$, we have from (B6), and Lemma 1 that

$$\overline{1 - \lambda} \approx 1 + \frac{1}{\sigma} n_c,$$

and
$$\lim_{\sigma\to\infty} \frac{d(1 - \lambda)}{dp}\Bigg|_{\frac{\pi_s}{1-\lambda}} = \lim_{\sigma\to\infty} \frac{(1 - F)^3}{c\cdot f\cdot \sigma} = \lim_{\sigma\to\infty} \frac{1}{c\cdot f\cdot\sigma},$$

or using the preceding relation, we obtain

$$\lim_{\sigma\to\infty} \frac{d(1 - \lambda)}{dp}\Bigg|_{\frac{\pi_s}{1-\lambda}} \approx \lim_{n_c\to-\infty} \frac{\bar{\lambda}}{c} \frac{1}{(-n_c)\cdot f(n_c)}.$$

But $\lim_{n_c\to-\infty} (-n_c)f(n_c) = 0$, so since $\lambda > 0$, it follows

that $\lim_{\sigma\to\infty} \frac{d(1 - \lambda)}{dp} = \infty.$ Q.E.D.

Just as in Chapter 2, we will want to obtain an idea of how large the customer per firm ratio, L/N, has to be before the deterministic analysis can even hope to meet desired tolerance limits. If the tolerance limits require $p < \bar{p}$ and $1 - \lambda < \overline{1 - \lambda}$, then we want to solve for the value of σ^2 (i.e., L/N) that satisfies

$$\overline{1 - \lambda} = 1 - \frac{1}{\sigma} H(n_c) \tag{B8}$$

where $$H(n_c) = f(n_c) - [1 - F(n_c)]n_c,$$

and $$1 - F(n_c) = c/\overline{p}.$$

The relation (B8) comes directly from (B7). Solving (B8) for σ^2 we obtain

$$\sigma^2 = \frac{[H(n_c)]^2}{\overline{\lambda}^2} \tag{B9}$$

To calculate the required σ^2 from (B9), we make use of the table below which calculates $H(n_c)$ as a function of p.

Table B-1

Price	$H(n_c(c/p))$
1.01 c	2.5
1.02 c	2.1
1.03 c	1.94
1.10 c	1.34
1.20 c	1.08
2.00 c	0.40

So, for example, if we allow the deterministic analysis to make only 1% errors in price and probability of shortage (i.e., $\overline{p} = 1.01 \cdot c$ and $\overline{\lambda} = .01$) then the minimum customer per firm ratio that is required before the deterministic analysis can hope to meet these tolerances is

$$\frac{(2.5)^2}{(.01)^2} = 62,500.$$

Allowing General Random Distributions for Demand

In Chapter 3, we examine the situations where price is fixed at $p\ (>c)$ and demand x equals $\bar{x} + e$, where e is a random variable with mean zero and density function $f(e)$. The expression for profits equals

$$\pi(S) = p\int_{-\infty}^{S-\bar{x}} (\bar{x} + e)f(e)d(e) + p \cdot S \int_{S-\bar{x}}^{\infty} f(e)de - cS,$$

where S = total stock. Differentiating the above to determine the optimal S, we obtain

$$\frac{\partial\pi}{\partial S} = 0 \Rightarrow p \cdot S\, f(S - \bar{x}) - p\, S\, f(S - \bar{x}) + p \int_{S-\bar{x}}^{\infty} f(e)de - c = 0,$$

or $\quad F(S - \bar{x}) = 1 - c/p,$

where $F(S - \bar{x}) \equiv \int_{-\infty}^{S-\bar{x}} f(e)de.$

Since $\quad \frac{\partial^2\pi}{\partial S^2} = -pf(s - \bar{x}) < 0$, the second order conditions are satisfied, so that the above first order condition defines an interior maximum.

CHAPTER 4

The Social Welfare Implications of Markets That Operate Under Uncertainty

4.0 Introduction

The previous chapters have examined how markets operate when the production decision must be made before the uncertain demand for the product can be observed, and when prices, once set, cannot vary over the market period. It is characteristic in such markets for some customers to be unable to purchase the good, and at the same time for firms to bear some risk that they will be unable to sell all of their goods. Equilibrium for such markets is described by the price for the good and by the probability of customer satisfaction. These two quantities, in turn, determine the total amount of the good that is demanded by customers, and the total amount supplied by firms. As seen earlier, the total amount supplied and demanded will not in general be equal in equilibrium. Depending on the assumptions about the information flows and institutional structure, three different equilibria were distinguished, the competitive equilibrium, the fair trade equilibrium, and the monopoly equilibrium. An important question to ask is whether any of the three equilibria involves a combination of price and probability of satisfaction that is optimal in the sense of maximizing some measure of social welfare. This is the issue that we examine in this chapter.

The first question we ask is when, if ever, will any of the previously discussed market equilibria maximize the expected value of the total consumer surplus. This question is motivated by two considerations. First, deterministic markets in competition maximize consumer surplus, so it is natural to see if uncertain markets do also. Second, expected consumer surplus is often used as an approximate measure of social welfare.[1] We will show that, in the special case where expected consumer surplus represents an individual's preferences between price and probability of satisfaction, the competitive equilibrium does indeed maximize the expected value of consumer surplus to society.

Consumer surplus as a measure of social welfare is known to suffer from several defects because of its partial equilibrium nature. Moreover, in an uncertain setting expected consumer surplus may not properly reflect consumer attitudes toward risk. To avoid these defects associated with consumer surplus, we examine the social welfare question in a simple two good model. We set up a two good model by introducing an alternative (non-rationed) good, and ask how a social planner who takes both markets into account would operate this economy so as to maximize the expected utility of a representative consumer. It will be shown that the socially optimal solution will usually be different from any of the previously discussed market equilibria. The socially optimal solution will, in general, involve paying lump sum subsidies to the firms that deal with the good that is subject to shortages. Compared to the social optimum, the competitive equilibrium will usually not devote sufficient resources to production of the good that is subject to shortages.

---

[1] See for example G. Brown Jr. and M. B. Johnson, "Public Utility Pricing and Output Under Risk," American Economic Review, 1969.

4.1 Maximizing Expected Consumer Surplus

As mentioned above, consumer surplus is not generally a good measure of social welfare for uncertain markets because, aside from the known problems associated with its partial equilibrium nature, it may not reflect consumer preferences between the probability of obtaining the good and the price of the good. For the special case where expected consumer surplus does reflect consumer attitudes toward risk, we want to examine whether any of the three previously examined market equilibria maximize expected consumer surplus. The main result of this section is that for this special case, the competitive equilibrium does maximize the expected consumer surplus to society.

The model is the same as before. There are N identical firms and L identical consumers with per capita demand x(p). Consumers randomly frequent one firm each period, and if the good is available, they purchase it according to their demand schedule. The cost to the firm for producing the perishable good is c per unit. At the beginning of each market period, firms must decide what price to charge for the good for that market period, and how much of the good to stock, or equivalently, how many customers, s, to be able to service for that market period. No deliveries of the good can occur once a market period has begun. Goods unsold at the end of the market period are discarded.

Let us consider the expression for expected consumer surplus to society when all N firms follow the same price and stocking policy. Expected consumer surplus to society equals the per capita consumer surplus times the number of customers times the expected fraction of customers that are satisfied minus the cost of the goods. Expressed mathematically, we have that

$$\text{Expected Consumer Surplus to Society} \equiv CSS = (1 - \lambda(s)) \int_0^{x(p)} x^{-1}(q)dq \cdot L - c\,s\,N\,x(p),$$

where
$$1 - \lambda(s) = \frac{L}{N} I(s) + s/L/N, \qquad (1)$$

and where

L = number of customers in the market,

N = number of firms,

$x(q)$ = the per capita demand curve,

$x^{-1}(p)$ = the inverse per capita demand function,

$1 - \lambda$ = the probability of satisfaction

$I(s) = \int_{-\infty}^{u} (t - u)\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} t^2} dt$, where $u \equiv (s - L/N)/(L/N)^{1/2}$,

p = price of the good,

c = cost per unit of the good, and

s = the number of customers that can be serviced per firm.

The relation (1) was derived earlier in Appendix A, and expresses the relation between the probability of stisfaction and the customer capacity, s.

The government wishes to determine an operating policy, (i.e. s and p), so that expected consumer surplus to society (CSS) is maximized when all firms behave according to this operating policy. To maximize CSS with respect to s and p, substitute (1) into CSS and take derivatives to obtain the following first order conditions:

$$(1 - \lambda)\, x'(p)\, p\, L - c\, s\, N\, x'(p) = 0,$$

or equivalently[1]

$$(1 - \lambda)\, p L - c\, s N = 0 \qquad (2),$$

and

$$(1 - F(u))\frac{L}{L/N} \int_0^{x(p)} x^{-1}(q)dq - c\, x(p)\, N = 0$$

where F(u) is defined as $F((s - L/N)/(L/N)^{1/2})$ and F(u) is the cumulative normal distribution. This last expression can be rewritten as

$$(1 - F(u))\, \frac{\int_0^{x(p)} x^{-1}(q)dq}{x(p)} - c = 0 \qquad (3).$$

Equations (2) and (3) determine the s and p of the operating policy for each firm that the government should follow to maximize the expected total consumer surplus to society.[2] Using the expression for profits derived in Chapter 2 (i.e. equations (2) and (4) of Section 2.3 and 2.6.3, respectively of Chapter 2) it can be seen that (2) is equivalent to the condition that profits per firm equal zero. Equation (3) determines the point along the zero profit (i.e. the $\pi = 0$) curve at which the government should operate.

Since profits equal zero at the point that maximizes consumer surplus to society, the only previously examined market structure that could produce this point is the competitive case discussed in Chapter 2. Even in this competitive case, the market equilibrium will not necessarily occur at the point that maximizes consumer surplus because the expected value of consumer surplus may not reflect the preferences of consumers toward risk.

---

[1]We assume $x'(p) < 0$.

[2]As in Chapter 2, we disregard the uninteresting case of boundary solutions (see Section 2.6.2 for a fuller discussion) and assume that (2) and (3) have a solution that represents an interior maximum.

The question then arises as to whether the competitive market equilibrium would maximize the expected value of consumer surplus to society if consumers' tradeoffs between the price of the good and the probability of obtaining the good were adequately represented by the expected value of their consumer surplus. At first glance, the answer to this question appears obvious. If expected consumer surplus reflects consumer preferences, then we know that the expected consumer surplus of each individual is maximized in the competitive market. Hence, the social planner will maximize the consumer surplus to society at this point. This reasoning is faulty although the conclusion turns out to be correct. The sum of individual consumer surplus does not equal the consumer surplus to society for the markets under study. There are unsold goods at the end of each period which must enter the government's calculation of consumer surplus but not that of any individual.

More specifically, suppose each of the L consumers seeks to maximize

$$\text{Consumer Surplus to an Individual} \equiv CSI = (1-\lambda)\left[\int_0^{x(p)} x^{-1}(q)dq - p\,x(p)\right] \quad (4)$$

where the notation was defined previously beneath (1).[1] Summing CSI over all L consumers and comparing this sum to the objective function, CSS, of the government, we see that the two expressions differ by

$$(1-\lambda)pLx(p) - csx(p)N = x(p)[p(1-\lambda)L - sNc].$$

---

[1]Consumers will maximize CSI if their Von Neumann utility functions are of the form $u(x_1,x_2) = g(x_1) + x_2$, where x_1 = the good under analysis, and x_2 = all other goods. The reader familiar with the public finance literature on option demand and consumer surplus should see the relation between that discussion and (4). See, for example, "Option Demand and Consumer Surplus: Further Comment" by C. J. Cichetti and A. M. Freeman III, Quarterly Journal of Economics, 1971, p. 536 footnote 7.

This last expression is the difference between the expected revenue to be received and the cost of all the goods, sold and unsold. In view of the differences in the objective functions between the individual and the government, it is interesting that the following theorem holds.

<u>Theorem 1</u>: Suppose expected consumer surplus to an individual, CSI, as defined in (4), represents consumer preferences between the price, p, and, probability of satisfaction, $1 - \lambda$. Then, the competitive equilibrium discussed in Chapter 2 maximizes the expected value of consumer surplus to society (CSS).

<u>Proof</u>: If CSI reflects consumer preferences, then from Chapter 2, we know that the competitive equilibrium occurs at that point along the zero profit curve that maximizes CSI. From (2), we know that the point that maximizes CSS also occurs along the zero profit curve.

The difference between consumer surplus to society, CSS, and the sum of consumer surplus to an individual, $L \cdot CSI$, was derived above and equals $x(p)[(1 - \lambda)pL - N s c]$. However, from (2), we see that along the zero profit curve, this difference equals zero. Therefore, along the zero profit curve, the two measures, CSS and $L \cdot CSI$, attain their maximum values at the same point. Q.E.D.

We see then that if individual consumer preferences are represented by expected consumer surplus (CSI), then just as in deterministic markets, the competitive euqilibrium will maximize the expected value of the total consumer surplus to society(CSS). Notice that price exceeds c and firms earn zero expected profits when expected consumer surplus is maximized. These results contrast sharply with those of other models that appear in the public finance literature and deal with a similar type of problem.[1] The results of those other models imply that to maximize expected consumer surplus to society, price should in general be less than c, and hence firms should operate at an expected loss.

The reason for the difference between the results of this chapter and those of other models lies in the manner in which the randomness is introduced into the demand curve. In the model of this chapter, a firm's demand is multiplicative and equals $x(p) \cdot i$ where $x(p)$ equals per capita demand and i equals the random number of consumers who visit the firm. In the other models in the public finance literature, the randomness enters additively, not multiplicatively, so that demand equals $x(p) + \varepsilon$, where $x(p)$ equals expected aggregate demand and ε is a random term. Additive demand implies that the absolute variation in demand is independent of the level of expected demand. So, for example, a 100 unit demand deviation is regarded as equally likely whether expected aggregate demand is 200 or 2 million. Multiplicative uncertainty implies that the relative variation in demand is independent of the level of expected demand. So, for example, a 1% deviation in demand is regarded as equally likely whether expected aggregate demand is 200 or 2 million. For most purposes, the multiplicative formulation would appear more plausible.[2]

If expected consumer surplus, CSI, does not represent consumer preferences for the probability of satisfaction, $1 - \lambda$, and the price, p, then Theorem 1 will not hold. However, if CSI does not represent consumers'

---

[1] G. Brown Jr. and M.B. Johnson, op. cit., and M.L. Visscher, "Welfare-Maximizing Price and Output with Stochastic Demand: Comment," American Economic Review, 1971.

[2] This is one reason why econometric equations are specified so often in log-log form.

preferences toward risk, then the expected consumer surplus is a very poor criteria to use as a measure of market performance in an uncertain environment.[1] Moreover, as mentioned earlier, consumer surplus is a partial equilibrium concept that ignores preferences for other goods, while the social optimum should take into account consumer preferences for other goods.[2] In the next section, we allow the consumer to have quite general preferences for the probability of satisfaction and the price, and examine how the introduction of an alternative good affects the anlaysis of the social optimum.

4.2 The Social Optimum in a Simple Two Good Model

Let there be two goods on which each of the L consumers can spend their identical endowment Y. Good 1 is the good that is subject to shortages. Good 2 is always available from the outside world at a constant price. The price of good 1 is p, while the price of good 2 is 1. As before, each unit of good 1 uses up c units of resources and must be produced before any firm observes its random demand. Demand is random in the same manner as discussed previously. As usual, no firm can receive delivery of the good once a market period has begun. The government owns each of the N firms that dispense good 1, and wishes to choose the same tax policy and operating policy for each of the N firms so as to maximize the expected utility of a representative consumer. The government faces the budget constraint that the sum of the firms' expected profits plus the total taxes collected or dispersed must equal zero.

[1] This point is not addressed by either Brown and Johnson, op. cit., or Visscher, op. cit.

[2] The only case in which it is not misleading to ignore other goods, is when the utility function is of the special form $g(x_1) + x_2$, where x_2 represents other goods.

Let u(x,z) represent the Von Neumann utility function of each consumer where x denotes good 1 and z denotes good 2. When good 1 is obtainable at price p, the utility of each consumer is given by V(p,Y), the indirect utility function. When good 1 is not obtainable, the utility of each consumer is given by u(o,y). If $1 - \lambda$ represents the probability of obtaining good 1, then the expected utility of a representative consumer can be written as

$$U(1 - \lambda, p) = (1 - \lambda)V(p,Y) + \lambda u(0,Y).$$

The government seeks to determine a transfer, T, for each individual,[1] a price, p, and a customer capacity s (recall that s refers to the maximum number of customers that can be serviced at any firm in any market period), so that the expected utility of any consumer is maximized. The government's problem can be written as

$$\max_{s,p,T} \quad (1 - \lambda(s))V(p,Y + T) + \lambda(s)u(0,Y + T) \tag{5}$$

subject to the budget constraint,

$$\pi(s,p) - \frac{L}{N} T = 0, \tag{6}$$

where $\pi(s,p)$ is the expression for expected profits per firm, which can be written as (using equations (2) and (4) of Sections 2.3 and 2.6.3, respectively, of Chapter 2)

$$\pi(s,p) = (1 - \lambda(s))p \frac{L}{N} x\,(p) - c\,s\,x(p) \tag{7}$$

where $1 - \lambda(s)$ is the expression for the probability of satisfaction as a function of customer capacity, s, and is given by (1).

[1]The variable T is the transfer from the firms to each consumer. Hence, if T < 0, consumers pay a tax, while firms receive a subsidy.

From the statement of the problem, we see that if (and only if) the transfer, T, equals 0 in the socially optimal solution, then it follows that the competitive equilibrium will also be the socially optimal point since both points maximize expected utility subject to the constraint that expected profits are zero. In general, there is no reason to expect that the optimal solution to the above problem will have T = 0, so that the competitive equilibrium will usually not represent the social optimum. The social optimum will usually involve either taxes or subsidies for the firms who sell good 1, the good subject to shortages. In such cases government intervention into a competitive market will be called for.

In order to investigate the conditions under which either taxes or subsidies will be paid in the social optimum, it is necessary to make an assumption about consumers' preferences.

Assumption 1: The marginal utility of an extra dollar, when good 1 is obtainable, is higher than the corresponding marginal utility when good 1 is unobtainable. More precisely, $V_2(p,Y) > u_2(0,Y)$ for all, p,Y, where the subscripts denote partial differentiation.

The assumption reflects the idea that the greater the variety of goods that can be purchased, the higher is the marginal utility of an extra dollar. (One sufficient condition for this assumption is that $u_{21} \geq 0$.) Given the above assumption, the following theorem holds.

Theorem 2: Under Assumption 1 and the assumption that per capita demand depends positively on income, the social optimum involves operating the N firms that sell good 1 at a loss and using lump sum taxes to subsidize their operation.

Proof: The Lagrangian for the government's maximization problem can be written as

$$\mathcal{L}(s,p,T,\mu) = (1 - \lambda)V(p,Y + T) + \lambda u(0,Y + T)$$

$$- \mu[((1 - \lambda)p \cdot \frac{L}{N} - cs)x(p,Y + T) - \frac{L}{N} T], \quad (8)$$

where μ is a Lagrange multiplier.[1] The first order conditions are:[2]

$$(1 - \lambda)V_1 = \mu[(1 - \lambda) \frac{L}{N} x + [(1 - \lambda)\frac{L}{N} p - sc]x_1], \quad (9)$$

$$(1 - \lambda)V_2 + \lambda U_2 = \mu[(1 - \lambda) \frac{L}{N} p - sc]x_2 - \frac{L}{N}], \quad (10)$$

$$(1 - F)\frac{N}{L}[V - U] = \mu[(1 - F)p - c]x, \quad (11) \text{ and}$$

$$[(1 - \lambda)\frac{L}{N} p - cs]x = \frac{L}{N} T \quad (12)$$

Substituting (12) into (9) and (10), we obtain

$$(1 - \lambda)V_1 = \mu[(1 - \lambda) \frac{L}{N} \cdot x + \frac{L}{N} \frac{T}{x} x_1] \quad (13)$$

and $(1 - \lambda)V_2 + \lambda U_2 = \mu[\frac{L}{N} \frac{T}{x} x_2 - \frac{L}{N}].$ (14)

---

[1]The notation was defined previously. Recall that λ is not a Lagrange multiplier, but is the probability of disappointment which is a function of s given in (1).

[2]Subscripts denote partial differentiation.

Since V is an indirect utility function, we have that $x = -\frac{V_1}{V_2}$. Using this relation, rewrite (13) as

$$(1 - \lambda)V_1 = \mu[(1 - \lambda)\frac{L}{N} + \frac{L}{N}\frac{T}{x}\frac{x_1}{x}] \cdot (-\frac{V_1}{V_2}),$$

or $$(1 - \lambda)V_2 = (-\mu)[(1 - \lambda)\frac{L}{N} + \frac{L}{N}\frac{T}{X} \cdot \frac{x_1}{x}],$$

or, $$V_2 = (-\mu)[\frac{L}{N} + \frac{L}{N}\frac{T}{x} \cdot \frac{x_1}{x}\frac{1}{1 - \lambda}]. \tag{15}$$

From (14) and Assumption (1), it follows that

$$\mu[\frac{L}{N}\frac{T}{x}x_2 - \frac{L}{N}] < V_2 \tag{16}$$

Substituting the expression for V_2 from (15) into (16), we have that

$\mu[\frac{L}{N}\frac{T}{x}x_2 - \frac{L}{N}] < (-\mu)[\frac{L}{N} + \frac{L}{N}\frac{T}{x}\frac{x_1}{x}\frac{1}{1 - \lambda}]$, or

$(-1)(-\mu)\frac{L}{N}\frac{T}{x}x_2 - \mu\frac{L}{N} < -\mu\frac{L}{N} - \mu\frac{L}{N}\ \frac{T}{x}\frac{x_1}{x}\frac{1}{1 - \lambda}$, or since $-\mu > 0$,

$$(-1)\frac{Tx_2}{x} < \frac{T}{x}\frac{x_1}{x}\frac{1}{1 - \lambda}, \text{ or}$$

$(-1)T\, x_2(1 - \lambda) < T\, x_2(1 - \lambda)$, or

$$T(-\frac{x_1}{x \cdot x_2}) < T(1 - \lambda). \tag{17}$$

If $T > 0$, then $\frac{-x_2}{x \cdot x_2} < 1 - \lambda < 1$, while if $T > 0$, then $\frac{-x_1}{x \cdot x_2} > 1 - \lambda$. But from the Slutsky equation, we know that $x_1 + x \cdot x_2 < 0$ or $\frac{-x_1}{x \cdot x_2} > 1$. Therefore if $T > 0$, we obtain a contradiction. Hence only $T < 0$ is possible in the optimal solution. Q.E.D.

Therefore, under Assumption 1, the socially optimal solution involves operating the N firms that produce good 1 at a loss, and using lump sum taxes to subsidize these firms' revenues. Since none of the previously discussed equilibria involve negative profit rates for firms and lump sum transfers from consumers to firms, it follows that none of these possible market equilibria will ever lead to the socially optimal point. Government intervention will be necessary to achieve the social optimum.[1]

The heuristic reason why under Assumption 1, it is optimal to tax consumers and pay subsidies to firms can be seen as follows. There are two states in which the consumer can wind up, one where he can purchase the good at the market price and one where he cannot. Under Assumption 1, the last dollar is more valuable in the state in which the good is obtainable than in the state in which the good is unobtainable. A person could increase his utility if he could in some way transfer part of his income between the two possible states. Such transfers of income are impossible in the problem under examination. However, what is possible is that the government can use taxes to reduce the income of a consumer in both states, and subsidize the operation of firms that produce the good and thereby reduce the price of the good subject to shortages. In this way, a transfer of purchasing power can occur between the two possible states in which the consumer can find himself. It turns out that this price reduction is always sufficient to overwhelm the decline in income, so that under Assumption 1 imposing some taxes always raises expected consumer utility.

[1]As should be clear from the proof of the theorem, if we replace Assumption 1 with the (less plausible) assumption that the marginal utility of income declines as the variety of goods increases, then in the social optimum firms would be taxed and consumers subsidized. In this case also, government intervention will be required to achieve the social optimum because (even if the consumers own the firms and share equally in the profits) there is no reason to expect any of the previously discussed equilibria to lead to the social optimum.

Theorem 2 tells us that the competitive equilibrium will not achieve the social optimum. Can we say whether, under Assumption 1, the competitive equilibrium will devote too few resources to firms selling good 1 and/or will involve a higher price for the good than occurs in the social optimum? Without further restrictions, all that can be said is that in the social optimum either the probability of satisfaction, (or equivalently the customer capacity, s) will be higher and/or the price of good 1 will be lower[1] than in the competitive equilibrium. We expect the normal case to involve an increase in the probability of satisfaction $1 - \lambda$, and a decrease in the price p. For this normal case it immediately follows that under Assumption 1 the competitive equilibrium will involve devoting too few resources (i.e. c s N x(p)) to the production of the good that is subject to shortages, when compared to the social optimum.

4.3 Summary

This chapter has examined the social welfare implications of markets characterized by demand uncertainty, price inflexibility and a lead time required for production. In the special case where expected consumer surplus reflects consumer preferences between the price of the good and the probability of obtaining it, it was found that the competitive equilibrium coincides with the social optimum since for this special case the

[1] In fact, in the social optimum, no only is it possible for the price to be below c, but it is even possible for the price to fall to zero. For example, if $V(p,Y) = \frac{Y}{p}$ and $u(0,Y) = 0$, then the social optimum involves the price falling to zero and the transfer going to -Y in such a way that $\frac{(Y + T)}{p}$ remains finite.

competitive equilibrium maximizes the total expected consumer surplus to society. However, there is no reason in general to believe that expected consumer surplus will properly reflect consumer preferences between price and the probability of obtaining the good. Moreover, as is well known, consumer surplus as a measure of social welfare suffers from several defects because of its partial equilibrium nature. To avoid these defects, we set up a two good model and asked how a social planner, who takes both markets into account, would operate the economy so as to maximize the expected utility of a representative consumer. It was found that in general none of the previously examined market equilibria would achieve the social optimum. Government intervention is required to achieve the social optimum.

By making a plausible assumption about how the marginal utility of income behaves as the variety of goods increased, it was possible to derive the properties of the social optimum. In the social optimum, consumers are taxed and firms that produce the good subject to shortages receive subsidies which enable the firms to offer more desirable price-probability of satisfaction combinations. This tax-subsidy scheme is able to transfer purchasing power from the state in which the consumer is unable to obtain the good to the state in which the consumer is able to purchase the good. In the social optimum more resources will usually be devoted to the production of the good that is subject to shortages than in the competitive equilibrium.

CHAPTER 5

Firm Interaction Under Uncertainty and Its Relation to Vertical Integration

5.0 Introduction

The previous chapters have examined how a market characterized by random demand, price inflexibility and a lead time required for production operates. The purpose of this chapter is to investigate how firms, in different markets, interact and affect each others behavior. In particular, this chapter will focus on the incentives for vertical integration caused by the transmission of uncertainty between a final product market and one of its factor markets. Firms will be assumed to compete with each other, and markets to reach equilibrium in the manner described in Chapter 2.

In the models of competitive markets that economists usually study, the very notion of shortage never appears, and hence the concern with an assurance of supply cannot even be understood. In sharp contrast, for the markets under study here, it is natural to have a firm occasionally be unable to satisfy some of its factor demand. The need for assurance of supply becomes an obvious and crucual variable in any firm's operating decisions. Focusing on the effects of the transmission of uncertainty between firms, we will show that it is socially undesirable to allow vertical integration to occur in these situations, yet that there will exist strong private incentives for at least some vertical integration to occur for every firm. The public policy implications regarding government policy in the markets under study would be to prevent such vertical integration from occurring.

Two types of vertical integration will be possible, complete and partial vertical integration. In the market structure involving complete vertical integration, the factor markets disappear, and each final product firm produces its own inputs. For the case of complete integration, it will be shown that, in general, the price of the final product is higher, and that more customers face shortages than in the more socially preferred case when vertical integration is not allowed. For the market structure involving partial vertical integration, a final product firm finds it cheaper to satisfy some of its uncertain demand for its product by using inputs purchased in the factor market, rather than relying exclusively on its own production of the input. For this case, it will be shown that the factor market acts as a type of insurance market for the final product firms. A final product firm is charged a premium to purchase the input in the factor market and will earn less net revenue when it sells a final product made with a factor market input than when it sells a final product made with an input that it produces for itself. In these partially integrated markets, the price of the factor input will usually be higher than it would be if vertical integration were not allowed.

The effect of the transmission of uncertainty between markets on the choice of technology will be examined. It turns out that a market structure involving no vertical integration could inhibit the introduction of a desirable new technology, while a structure involving complete vertical integration could lead to the rapid introduction of such a technology.

Finally, we will examine within the context of uncertain markets whether vertical integration through internal growth is more desirable than vertical integration through the acquisition of firms. Focusing on the effects of uncertainty, we will find that internal growth can be a much more harmful method of vertical integration than the acquisition of firms. This result contrasts with the usual discussions of antitrust policy where, based on considerations of market power, the reasoning is that vertical integration through internal growth is acceptable, while vertical integration through acquisition is not.

This chapter is organized as follows. Section 5.1 discusses the previous attempts in the literature to deal with the question of vertical integration. Section 5.2 presents a model of the transmission of uncertainty in a market structure where demand uncertainty and price inflexibility is present at both the final product and input market levels. Sections 5.3 through 5.8 draws on the theorems and proofs of the previous chapters to establish results about the incentives and consequences of vertical integration for the markets under study. Section 5.9 discusses the effect of the transmission of uncertainty on the choice of technology. The remaining sections deal with the issue of internal growth versus merger and with applications of the model to questions other than vertical integration.

5.1 Literature on Vertical Integration

"The study of vertical integration has presented difficulties at both the theoretical and policy levels of analysis. That vertical integration has never enjoyed a secure place in value theory is attributable to the fact that, under conventional assumptions, it is an anomaly."[1] In a world of pure competition in both the input and final product markets, with constant returns to scale, there is absolutely no incentive to vertically integrate. Students of market organization have put forth a number of situations where the conventional assumptions break down and an incentive for vertical integration exists. Econo-

[1] O. Williamson - "The Vertical Integration of Production: Market Failure Considerations", American Economic Review, May, 1971.

mies of scale in production, monopoly power in either the final product or input market, tax considerations, informational advantages or imperfections in the factor market are all justifications for vertical integration. It is only this last reason that will possibly bear a relation to the issues that are discussed in this chapter.

The literature in industrial organization is replete with statements to the effect that it is the uncertainty of factor supplies that creates incentives for vertical integration. For example, Chandler, in his discussion of the reasons for the formation of the largest companies in the United States argues "the initial motives for expansion or combination and vertical integration had not been specifically to lower unit costs or to assure a larger output per worker by efficient administration of the enlarged resources of the enterprise. The strategy of expansion had come...from the desire...to have a more certain supply of stocks, raw materials and other supplies..."[1] In studying Dupont's reasons for integrating, Chandler finds that "the need for assured supplies demanded increasing vertical integration."[2] Regarding General Motors, we find that "Durant personally organized a number of them [i.e. vertically integrated] in order to make certain that his assembly line would have a dependable supply of parts."[3] Despite the frequency with which the argument about the need for assured supplies appears in historical studies of vertical integration, it is usually never explained very well why the factor supply is uncertain, or why an uncertain factor supply should create incentives for vertical integration.

[1] A. Chandler, Jr. - Strategy and Structure: Chapters in the History of American Industrial Experience, M.I.T. Press, 1964, pp. 37.

[2] ibid. p. 84

[3] ibid. p. 116

At a more theoretical level, several authors concerned with industrial organization have suggested that uncertainty could provide an incentive for vertical integration. As early as 1937, Coase[1] argued that with constant returns to scale in production the very existence of a firm depends on some sort of market imperfections. The firm organizes when its internal allocative ability is superior to that of a market. Coase claimed that uncertainty about finding sellers of factors of production could provide one justification for the existence of a firm. Applying Coase's reasoning to the question of vertical integration many years later, Malmgrem[2] indicated that the presence of uncertainty could create incentives for vertical integration. "Activities which tended to fluctuate, causing fluctuations in prices and outputs in the market, could be integrated and balanced against one another."[3] Malmgrem argued that when prices do not reflect scarcity, vertical integration can occur. More recently, Williamson[4] discussed how uncertainty can make it difficult to establish contracts, and could provide incentives for vertical integration. Underlying all of the above arguments is the notion that uncertainty can somehow cause a breakdown in the functioning of a market. None of the discussions ever address the issue of whether vertical integration is a socially desirable response to the uncertainty in the market.

Until very recently, there had been no attempt to analytically investigate the claims of the above authors as regards the effect of uncertainty on the incentives for vertical integration. Within the past year, two economic

---

[1]R. Coase - "The Nature of the Firm," Economica, 1937.

[2]H. Malmgrem - "Information, Expectations, and the Theory of the Firm," Quarterly Journal of Economics, November, 1971.

[3]ibid.

[4]Williamson, op. cit.

theorists have sought to bridge the gap in the literature. Arrow[1] has shown that as in the deterministic case the presence of uncertainty in a factor market with a freely fluctuating price does not create any incentives for vertical integration. Arrow then proceeds to investigate the case where informational advantages accrue to vertically integrated firms. In a paper closely related in topic to this one, Green[2] showed that if rationing is possible in the factor market, then, unless we assume a priori that integrated firms are less efficient producers of the intermediate good than nonintegrated producers, every firm will have an incentive to fully integrate. In order to avoid the trivial solution where every firm has an incentive to fully integrate, Green postulates and examines the case when the integrated firm has an inferior technology for producing the input. (Green mentions in his footnote 10 that this assumption of an inferior technology for the integrated firm can be questioned.) Prices are exogeneous in Green's model so that there is an input market failure in the sense in which Malmgrem argued, in that prices need bear no relation to scarcity or rationing probability. In Green's model, final product firms face no uncertainty in their demand for their final product and are able to sell all of their product at the exogenous market price.

The focus of this chapter is to show that in markets characterized by the type of uncertainty and price inflexibility discussed in previous chapters, the transmission of uncertainty from the product market to the factor market can create socially undesirable incentives for vertical integration. In order to concentrate on the effects of the transmission of uncertainty, we will avoid making any of the traditional assumptions that lead to vertical

---

[1] K. Arrow - "Vertical Integration and Communication," Institute for Mathematical Studies in the Social Sciences, 1974. A condensed version of this paper appears in <u>The Bell Journal of Economics</u>, Spring, 1975.

[2] J. Green - "Vertical Integration and Assurance of Markets," Discussion Paper 383, Harvard Institute of Economic Research, 1974.

integration. In both the final product and factor market, firms will compete with each other, will be untaxed, and will have constant returns to scale in production. Integrated and nonintegrated firms will have the identical production technologies available to them so that there is no asymmetry in the production efficiency of the factor input. The prices in both the final and factor market will be endogeneously determined in accordance with the analysis of the previous chapters on how such markets operate. Therefore, prices will reflect the economic scarcity and rationing probability of goods. Firms will have an incentive to integrate to lower the probability of being unable to obtain the factor input (i.e., to better "assure" themselves of supplies.) Firms will have an incentive not to integrate to avoid the probability of being left with unused input.

5.2 The Model

This section presents a simple model of the transmission of uncertainty between a product market and one of its factor markets.

There are two types of firms, stage 1 and stage 2 type firms. Stage 1 firms require factor inputs from stage 2 firms to produce the final good. There are N_1 identical stage 1 firms and N_2 identical stage 2 firms, with N_2 less than N_1. Demand facing an individual stage 1 firm is random during any market period. Therefore, the derived demand for factor inputs from any stage 1 firm is random. In the stage 2 factor market, the demand facing any firm is also random. The final good cannot be produced without the factor input, and the amount of the factor input available in any period must be determined before any of the demands for the final product can be observed. Therefore, there is a risk that a unit of input will not be used by the time the market period ends. We assume that unused input is discarded at the end of the market period. However, even if inventory can be held from one period to the next, as long as there are costs to holding inventories, the same types of qualitative results as developed in this chapter will hold. Prices are set at the beginning of each market period before any demands are observed, and are not allowed to vary within any market period.

Stage 2 firms can deliver the input to stage 1 firms within any market period. We allow stage 1 firms the option of producing some of the factor input for itself. We refer to the production and holding of the input by stage 1 firms as vertical integration. If a stage 1 firm produces the factor input for itself, it bears the risk of having unused input at the end of the market period. A stage 1 firm is not allowed to sell its inputs in the stage 2 factor market. Stage 1 firms cannot ship the factor input between themselves, nor can stage 2 firms.

We assume that the production technologies for producing the final product and factor input both involve constant returns to scale. The same technology for producing the factor input is available to both stage 1 and stage 2 firms. As in the earlier chapters, it costs c to produce one unit of the factor input. The final product is produced by a Leontief technology that requires K units of capital and 1 unit of the factor input sold in the stage 2 market to produce one unit of the final good. The capital input is always available at a constant price r per unit.

The market operates as follows. As in the earlier chapters, there are L identical customers, each with a per capita demand curve x(p). In each market period, each of the L customers randomly frequents one stage 1 firm where he demands the final product. Every time a stage 1 firm observes a customer demand for its product, it attempts to obtain the factor inputs necessary to produce the customer's demand for the final product. The stage 1 firm first tries to use up its own holdings, if any, of the factor input, and then, when its factor holdings are depleted, it enters the stage 2 factor market. Once in the factor market, the stage 1 firm randomly frequents a stage 2 firm to try to obtain the necessary inputs to be able to satisfy its customer. If the stage 1 firm is unable to obtain the input from the stage 2 firm, then the stage 1 firm is unable to satisfy the demand of the customer. This customer returns home dissatisfied. As in Chapter 2, customers have preferences, which firms recognize, between the price of the good and the probability of obtaining that good. For any given level of factor holding by the stage 1 firms, we can imagine the stage 1 and stage 2 firms competing in their respective markets on the price and probability of satisfaction until each market reaches the competitive equilibrium described in Chapter 2.

The important feature of this market structure is that the amount of the factor input that stage 1 firms decide to hold affects both the level and the uncertainty of the demand that stage 2 firms see. If stage 1 firms hold none of the factor so that there is no vertical integration, then the stage 2 firms essentially face a random equal share process where the total number of customers equals L.[1] Whenever stage 1 firms hold some of the factor input for themselves, the total number of customers (i.e., visits from stage 1 firms) that all stage 2 firms see becomes a random variable whose expectation falls below its value of L in the no integration case. In accordance with the discussion in Section 2.7.2 of Chapter 2, we assume that in this case, the stage 2 firms behave as if they see a random equal share process, regarding

[1] Stage 1 firms, on behalf of their customers, enter the stage 2 markets once for each of the L customers. Also, recall that a random equal share process consists of each of the L customers randomly frequenting one of the N_2 stage 2 firms.

the total customer size as the expected number of customers who enter the stage 2 market.

5.3 Issues Associated with Firm Interaction Under Uncertainty

Now that the model of the transmission of uncertainty between the stage 1 and stage 2 markets has been described, we can state more clearly the issues that we wish to examine. The decision of stage 2 firms about how much of the input to produce affects the probability that a stage 1 firm will be able to obtain the input and produce the final product. If a stage 1 firm becomes dissatisfied with the operating policy of stage 2 firms, the stage 1 firm can produce some of the input for itself, and itself bears the risk of having unsold input at the end of the market period. The decision of stage 1 firms to produce some of the input for themselves and bear the risk of having unsold input, affects the entire stochastic structure of demand that the stage 2 firms will see , and hence influences the equilibrium that is reached in the stage 2 market[1] which, in turn, influences the equilibrium that is reached in the stage 1 market.

Firms in each market compete amongst themselves until equilibrium is reached in the manner described in Chapter 2. Therefore, we know that in the equilibrium in both the stage 1 and stage 2 markets, each firms' operating policy reflects the preferences of its customers, and that the prices reflect the probability of obtaining the good (i.e., the probability of satisfaction of Chapter 2). The important question to ask is whether, under competition, firms are forced to take into full account the effect of

---

[1]Recall from Chapter 2, that the stochastic structure of demand affects the operating cost of firms.

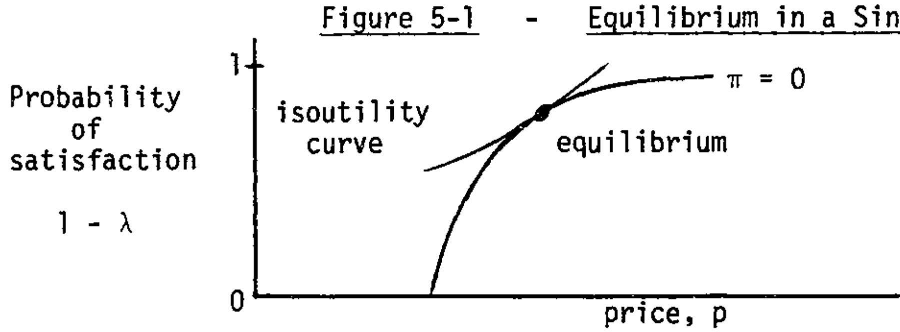
their operating policies on the transmission of uncertainty to other markets. What happens to the welfare of consumers as final product firms produce some of the factor input for themselves and themselves bear the risk of having unsold input? From the consumers' point of view, is there some preferred allocation between the stage 1 and stage 2 firms for producing the input and bearing the risk of having unsold inputs? Do the incentives under competition lead firms to adopt this preferred market structure, or, are the incentives under competition perverse, and discourage firms from adopting the preferred operating policies? If competition does not lead to the preferred market structure, how does the competitive equilibrium differ from the preferred one? If vertical integration is to occur, is it better to allow stage 1 firms produce the input for themselves, or is it better to have the N_1 stage 1 firms each try to acquire one of the N_2 stage 2 firms? (Obviously, $N_1 - N_2$ of the stage 1 firms will be unable to acquire a stage 2 firm and will be forced to go out of business.) If suddenly a new technique for producing the output becomes available, is it more likely to be adopted when the market structure is integrated or nonintegrated?

In the next section, we analyze in detail how the stage 1 and stage 2 markets reach equilibrium when firms in the same market compete with each other. The later sections draw on this analysis to answer the questions raised in this section.

5.4 Market Operation**

In this section, we describe how equilibrium is reached in the stage 1 and stage 2 markets, when each stage 1 firm produces for itself an amount of factor input so as to satisfy some given number of customers by itself. The analysis draws heavily on the discussion in Chapter 2 of how a single market subject to demand uncertainty, price inflexibility and a lead time required for production reaches equilibrium when firms compete amongst themselves. Let us briefly refresh the reader's mind of how such single markets equilibriate. Consumers have preferences between the price, p, of the good and the probability, $1 - \lambda$, of obtaining the good. Through competition, firms are forced to operate on the zero profit ($\pi = 0$) curve. The height of the zero profit curve depends on the number of customers, and the number of firms. Competitive equilibrium is defined as the point of tangency between the highest isoutility curve and the zero profit curve. This equilibrium is depicted below.

Figure 5-1 - Equilibrium in a Single Market



**This section is fairly technical and draws on much of the analyses of Chapter 2. The analysis of this section forms the basis of many of the formal proofs of the theorems of the upcoming sections. A reader not interested in the technical details of market operation can omit most of this section without loss of continuity, and still be able to understand the heuristic explanations of the theorems. The only part of this section that is essential if the reader is to understand the remainder of this chapter, is the last three paragraphs, where complete and partial vertical integration are defined.

Just as in Chapter 2 (see Section 2.6.2), we ignore the possibility of multiple tangencies and assume that in equilibrium, identical firms follow identical operating policies. As the diagram on the previous page illustrates, market equilibrium is characterized by a price and a probability of satisfaction.

Let me now describe in detail how the stage 1 and stage 2 markets operate. First, let me introduce some notation.

Let p_f = price of the final good;

p_{int} = price of the factor input, sold in the stage 2 market;

c = cost of producing the factor input;

$x(p_f)$ = per capita demand for the final product;

$1 - \lambda_2$ = probability of satisfaction in the stage 2 factor market;

$1 - \lambda_1$ = probability of satisfaction in the stage 1 market; when stage 1 firms rely only on themselves for input supplies;

$1 - \lambda$ = probability of satisfying a customer who demands the final product;

r = exogenous price of capital;

$(K, 1)$ = the input-output coefficients for producing the final product from capital and the stage 2 input;[1]

π_1 = profit of a stage 1 firm;

π_2 = profit of a stage 2 firm;

L = number of identical customers;

N_1 = number of stage 1 firms; and

N_2 = number of stage 2 firms.

The functional forms of the probability of satisfaction functions $(1 - \lambda_1)$ and the zero profit curves $(\pi_i = 0)$ were derived in detail in Appendix A. It

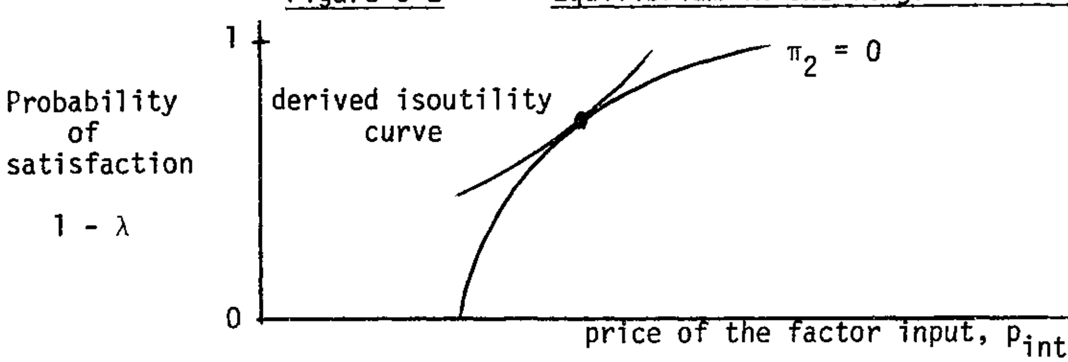
[1]The fact that there is capital will not play an important role in the model until we consider the choice of output technology in a later section. For that section, it is necessary to understand how the capital coefficient influences equilibrium. For this reason, we will develop the model in its most general terms, including capital in the model.

is important to remember that the $1 - \lambda_i$ functions and $\pi_i = 0$ curves each depend on the number of firms in the market, and on the total (expected) number of customers that visit the market each period. For emphasis, we shall sometimes write the $\pi_i = 0$ curve as $\pi_i(L_i^e, N_i) = 0$ where L_i^e and N_i are the expected number of customers and the number of firms in the stage i market.

Suppose, first, that stage 1 firms produce none of the input for themselves and rely entirely on the stage 2 firms for their input supply. Since both markets operate according to the principles described in Chapter 2, we know that in equilibrium stage 1 and stage 2 firms both earn zero profits. The zero profit condition for stage 1 firms, when they hold none of the factor input is equivalent to the condition that $p_f = rK + p_{int}$.

The isoutility curves of the consumers reflect the tradeoffs between the final price, p_f, and the probability of satisfaction, $1 - \lambda$, that leave consumers indifferent. When stage 1 firms enter the stage 2 factor market, they are willing to make any $(1 - \lambda, p_{int})$ tradeoff that would translate into a $(1 - \lambda, p_f)$ tradeoff that a consumer of the final good would make. Therefore, using the zero profit condition in the stage 1 market, and the isoutility curves of the consumer in $(1 - \lambda, p_{int})$ space, it is possible to derive what the isoutility curves look like in $(1 - \lambda, p_{int})$ space. As discussed in Chapter 2 the tangency between these "derived" isoutility curves[1] and the $\pi_2 = 0$

Figure 5-2 - Equilibrium in the Stage 2 Factor Market



[1]Note the similarity to deterministic theory where equilibrium in the factor market is determined by a derived demand curve. For the case under consideration here, equilibrium in the factor market is determined by the derived isoutility curves.

curve (i.e. the zero profit condition in the stage 2 factor market will establish equilibrium in the factor market. This equilibrium is depicted on the previous page.

Suppose that without the final good, the consumer can achieve a level of utility equal to u_o. Let u* be the level of utility achieved at the equilibrium point depicted in Figure 5-2. As discussed in Section 2.6.2 of Chapter 2, we will always assume u* is strictly greater than u_o in order to insure that the markets under examination exist. As long as consumer demand justifies the existence of the final product market then equilibrium in the stage 1 and stage 2 markets, in the case of no vertical integration by stage 1 firms, is completely determined by the point of tangency depicted in Figure 5-2, and by the zero profit condition in the stage 1 markets, namely $p_f = rK + p_{int}$.

Now suppose each stage 1 firm produces an amount of the factor input so that it is able to satisfy z customers by itself (i.e. customer capacity equals z). This amount of the factor input enables the stage 1 firm to satisfy $1 - \lambda_1(z)$, of its customers on average without relying on the factor market.[1] For this situation, the determination of the equilibrium conditions is more complicated than for the case when z = 0. Let us first consider the stage 1 firms.

The probability that a stage 1 firm will be unable to satisfy any customer will now depend on the probability of satisfaction in both the stage 1 and stage 2 markets. As a useful approximation, we have that the probability that a customer is satisfied = $1 - \lambda \approx 1 - \lambda_1\lambda_2$ where λ_i is the probability of disappointment in market i. In other words, if stage 1 firms, by themselves, can satisfy 80% of their customers, and stage 2 firms can satisfy 50% of their customers, then 90% of the total number of customers will be satisfied on average (80% + (100-80)% · 50%).

---

[1]The functional form of $1 - \lambda(z)$ was derived in Appendix A. For the interested
For the interested reader $1 - \lambda(z) = \frac{\sqrt{\frac{L}{N}}I(z) + z}{L/N}$ where $I(z) = \int_{-\infty}^{z} (t - z)e^{-\frac{1}{2}t^2} dt$.

Since stage 1 firms hold the factor input, the price, p_f, that they charge for their final product must now cover the expected costs incurred not only from producing with a stage 2 firm's input (cost = $p_{int} + rK$), but also from producing with their own input (cost = $c + rK$) and from having unsold input at the end of the market period. Using the notation introduced at the beginning of this section, we can write that profits for stage 1 firms with customer capacity z equals

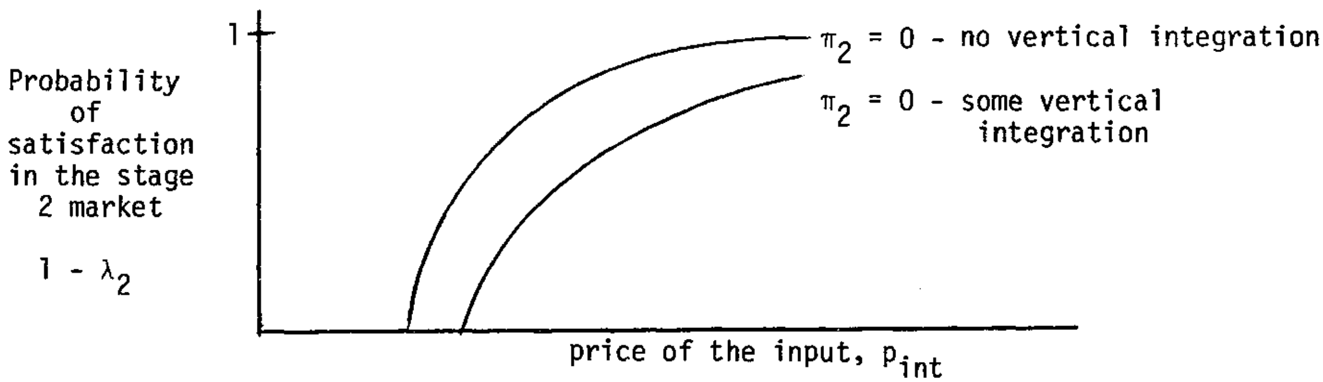
$$\pi_1(z) = [-cz + (1 - \lambda)p_f \cdot \frac{L}{N_1} - (1 - \lambda)\frac{L}{N_1}r \cdot K - \lambda_1 \cdot (1 - \lambda_2)\frac{L}{N_1}p_{int}]x(p) \quad (1)$$

where $\lambda = \lambda_1\lambda_2$ and $\lambda_1 = \lambda_1(z)$, and is given in the preceding footnote. The first term in the parenthesis in (1) is the input cost of having to produce the input $z \cdot x(p)$. The second term represents the expected revenue to be received. The third term is the expected capital cost to be incurred, while the fourth term is the expected cost to be incurred from input purchases in the stage 2 factor market. The zero profit condition for stage 1 markets requires that the expression in parenthesis on the right hand side of (1) be set equal to zero. Notice that when the stage 1 firms produce no input for themselves (i.e. $z = 0$), then the fraction of customers that they can satisfy themselves is zero (i.e. $\lambda_1 = 1$). In this case, the zero profit condition simply becomes $p_f = rK + p_{int}$, the condition that we used in the earlier discussion of market equilibrium when no vertical integration occurs.

Consider the stage 2 firms. Since some of the demand for the input is satisfied by stage 1 firms, the total number of visits that stage 2 firms will receive will fall from its value of L in the case involving no vertical integration. The stage 2 firms will only receive, in total $\lambda_1 \cdot L$ visits rather than L visits per period, when the input is held by stage 1 firms, where

$\lambda_1 = \lambda_1(z)$ is the probability that a stage 1 firm, with customer capacity z, will be unable to satisfy a customer by using its own inputs. This means that the zero profit curve in the stage 2 market (i.e. $\pi_2 = 0$) must now be derived under the assumption that the total number of customers is $\lambda_1 \cdot L$. Therefore, from the arguments given in Chapter 2, the new $\pi_2 = 0$ curve lies entirely below the $\pi_2 = 0$ curve in the case of no integration, as depicted below.

Figure 5-3 - Zero Profit Curves For Stage 2 Firms With and Without Vertical Integration



Essentially, as λ_1 falls from 1 (its value in the case of no integration), the stage 2 firms become less efficient absorbers of risk in the sense that stage 2 firms offer a worse "menu" of price-probability of satisfaction choices as stage 1 firms vertically integrate.

Consumers have preferences for the price of the final good, p_f, and the probability, $1 - \lambda$ of obtaining it. For any given customer capacity, z, that stage 1 firms provide for themselves, there is a relationship between the final probability of obtaining the good, $1 - \lambda$, and the probability of satisfaction, $1 - \lambda_2$, in the stage 2 market. The zero profit condition (1) in stage 1 markets establishes another relation between p_f, $1 - \lambda$, $1 - \lambda_2$, and the price of the stage 2 input p_{int}. Using these two relations, it is possible to construct "derived" isoutility curves in $(1 - \lambda, p_{int})$ space from the original isoutility curves in $(1 - \lambda, p_f)$ space. These derived isoutility curves reflect tradeoffs in the factor market that translate into tradeoffs in $(1 - \lambda, p_f)$ space which leave consumers indifferent. Stated more formally, given any z, with its associated λ_1, and $1 - \lambda_2$ and p_{int}, there exists a p_f from (1) such that profits are zero for the stage 1 firms. Furthermore, the final probability of satisfaction $1 - \lambda$ is given by $1 - \lambda_1\lambda_2$. Therefore, for any z, we can map each point in $(1 - \lambda_2, p_{int})$ space into a point in $(1 - \lambda, p_f)$ space, and vice-versa. Using this mapping we can construct derived isoutility curves in the factor market between $1 - \lambda_2$ and p_{int} that reflect consumer preferences for $1 - \lambda$ and p_f.

Given the level of customer capacity, z, of stage 1 firms, equilibrium can be derived in a fashion similar to that in the case of no integration. First, consider the stage 2 markets. Define the market clearing point in that market as the tangency in $(1 - \lambda_2, p_{int})$ space between the $\pi_2(\lambda_1 L, N_2) = 0$[1] curve and the highest derived isoutility curve. Let the level of utility associated with this tangency point be u*. The stage 1 zero profit condition (1), together with the relation $1 - \lambda = 1 - \lambda_1\lambda_2$, then determine the remaining market clearing quantities.

Now, suppose that the stage 2 firms disappear, so that the stage 1 firms can rely on only their own holdings of the factor input to satisfy customers. To determine market clearing in this situation, p_f would be chosen to yield zero profits, and $1 - \lambda$ would equal $1 - \lambda_1(z, N_1)$. Let the level of utility associated with this market clearing be u_0. If the level of utility when consumers do without the final product is not less than both u* and u_0, then neither the stage 1 nor stage 2 markets need exist. If this is not the case,

[1]The $\lambda_1 \cdot L$ and N_2 are written to emphasize that the $\pi_2 = 0$ curve is drawn under the assumption that N_2 stage 2 firms receive $\lambda_1(z) \cdot L$ visits, instead of L visits per period.

then there are two possible types of market equilibrium corresponding to whether $u^* \gtrless u_0$.

If $u_0 > u^*$, the market equilibrium, with stage 1 customer capacity z, will involve the disappearance of the stage 2 markets. A stage 1 firm will rely entirely on its own production of the factor input. This case, we call complete vertical integration.

If $u^* > u_0$, then market equilibrium involves operation of both the stage 1 and stage 2 markets, and is defined by the tangency point, discussed above, between the derived isoutility curves and the $\pi_2(\lambda_1(z)\cdot L, N_2) = 0$ curve. In this case, the stage 1 firm relies on both its own holdings and the stage 2 market to provide the necessary inputs for producing the final product. We call this case partial integration.

The reader may have noticed that in the case of partial integration, if a stage 1 firm has exhausted its own input holdings and receives a customer, the stage 1 firm will always enter the stage 2 market to try to secure the input for the customers. Market structures where a stage 1 firm flips a coin to decide to enter the stage 2 market are not considered. The appendix to this chapter, Appendix C, proves that such randomized strategies can never occur in equilibrium.

5.5 The Social Costs of Vertical Integration

The previous sections have shown that the production of the input by stage 1 firms will affect both the level and stochastic structure of demand facing stage 2 firms, and thereby influence the equilibrium that is reached in the stage 1 and stage 2 markets. In this section, the consequences of vertical integration on consumer welfare are examined. Let me reemphasize that the purpose of this chapter it to examine the effect of the transmission of uncertainty between interacting firms, and not to investigate all of the possible consequences of vertical integration. Other issues associated with vertical integration, such as the attainment of monopoly power or technological economies of scale, are purposely not modelled so as to focus solely on the effects of the transmission of uncertainty.

If all firms within any stage behave identically, then there is always a higher probability that an incremental unit of the factor input will be used if it is held in a stage 2 rather than a stage 1 firm. Stated in another way, since the number of stage 1 firms exceeds the number of stage 2 firms, a unit of the factor will be more frequently used if it is given to a stage 2, and not a stage 1, firm. From this simple observation, we can obtain the following:

Theorem 1: Any market structure involving vertical integration achieves a lower level of expected utility than does the market structure involving no vertical integration.

Proof:[1] Let the customer capacity of each stage 1 firm equal z (> 0). The stage 1 and stage 2 markets reach equilibrium in the manner described in the previous section. In equilibrium both stage 1 and stage 2 firms earn zero profits. Therefore, the total revenue net of capital cost taken in at stage 1 must pay for the costs of the total amount of the input held by both the stage 1 and stage 2 markets. Let $x(p_f^o)S^o$ = total amount of the factor input held by the stage 1 and 2 markets. Then, we can write this total revenue condition as $(1 - \lambda^o)L(p_f^o - rK) = c \cdot S^o$ where all variables with superscript

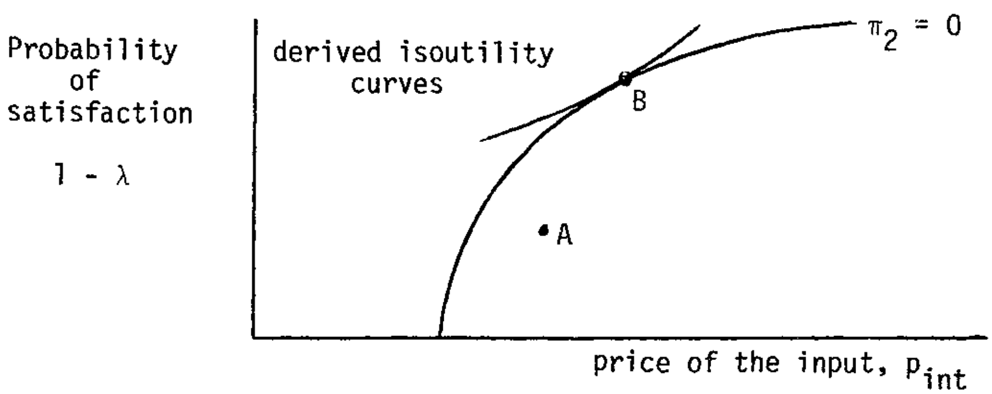
[1]This proof relies heavily on the analysis of the previous section. A heuristic explanation follows for the reader who omitted that section.

"o" refer to their values in equilibrium and were defined in Section 5.4. Consumers achieve a level of utility $u(1 - \lambda^o, p_f^o) = u_o$ in this market equilibrium.

Now, keep p_f^o fixed, but move all the stage 1 input holdings to stage 2 firms. By the above discussion, more consumers will be satisfied, so $1 - \lambda^o$ will rise to some $1 - \lambda^*$. But, then the expected net revenue, $(1 - \lambda^*)\cdot(p_f^o - rK)x(p_f^o)$, will exceed the total costs, $c\cdot S^o\cdot x(p_f^o)$ so that profits in the system will be positive. Alter the price of the factor input p_{int}^o to p_{int}^* so that profits in stage 1 markets are zero (i.e., set $p_{int}^* = p_f^o - rK$). Profits in the stage 1 market will now be zero, but profits in the stage 2 market will be positive since the system as a whole is taking in more revenue than it is paying out. Call the point $(1 - \lambda^*, p_f^o)$ point A. Since point A maps into $(1 - \lambda^*, p_f{}^\circ)$ with $1 - \lambda^* > 1 - \lambda^\circ$, the utility associated with point A, $u(1 - \lambda^*, p_f^o)$, exceeds u_o.

Consider equilibrium in the factor market in the case of no integration. This equilibrium is depicted below at point B as the tangency between the $\pi_2(L, N) = 0$ curve and the derived isoutility curves. Since point A involves positive profits to stage 2 firms it must lie below the $\pi_2(L, N) = 0$ curve. (See diagram.)

Figure 5-4 - Market Equilibrium With No Vertical Integration



It immediately follows then that point B, which represents equilibrium in the case of no integration, involves a higher level of utility than does point A which represents a level of utility higher than u_0. Hence the level of utility achieved in the equilibrium involving no vertical integration exceeds that achieved in the equilibrium involving vertical integration. QED

The reason why Theorem 1 is true can be explained intuitively as follows. The number of final product stage 1 firms exceeds that of factor input stage 2 firms. Therefore, stage 2 firms are more efficient absorbers of risk in the sense that stage 1 firms would have to hold more of the input than stage 2 firms in order to satisfy the same fraction of the population. Although holdings of the input by stage 1 firms reduces the demand seen by stage 2 firms, this reduction in demand is not great enough to offset the inefficient risk absorbtion by stage 1 firms. Therefore, stage 2 firms must decrease their input holdings by less than the amount that stage 1 firms increase their input holdings, if the same fraction of the population is to be satisfied. So, when stage 1 firms produce any input for themselves, more total input in the system must be produced or the fraction of customers who are satisfied will decline. Therefore, to satisfy any given fraction of consumers, market structures involving vertical integration will have higher input costs than those involving no vertical integration. Since competition insures that cost savings are passed on to consumers, it follows that consumers can always be made better off whenever there is any vertical integration in the system.

Theorem 1 can be heuristically explained in terms of sharing. Consider the following example. There are two bakeries, side by side, and 100 customers who each day randomly frequent one bakery and buy one loaf of bread. If the bakeries are willing to share their production of bread with each other, then only 100 loafs need be produced to satisfy the entire population. If the bakeries refuse to share (or else if it is costly to share) with each other, then each bakery must produce 100 loaves of bread to be sure to satisfy the customer population. Sharing allows 100, instead of 200, loaves to suffice. Exactly analogous reasoning applies to vertical integration. When stage 1 firms produce the input for themselves, they in effect do not share it with other stage 1 firms. However, when all stage 1 firms rely on stage 2 firms for the factor input, they are essentially "sharing" from common resource pools. Since insurance-like costs can always be lowered when no sharing is taking place, vertical integration can impose unnecessary costs on society.

So far, this chapter has: (1) described how a vertically integrated market structure would equilibrate in an uncertain market setting; and (2) shown that vertical integration can be socially undesirable. In the description of market operation, the customer capacity, z, of the stage 1 firms was always taken as given. The important questions regarding market equilibrium that remain are:

1. whether there are any incentives for vertical integration to occur at all,

2. if such incentives exist, what level of customer capacity z by the stage 1 firms will characterize the equilibrium market structure and

3. if such incentives exist, how will the vertically integrated equilibrium compare to the socially preferred non-integrated one.

Questions 1 and 2 are discussed in the next sections, while question 3 is dealt with in Section 5.8.

5.6 Private Incentives for Vertical Integration

Suppose that each stage 1 firm is holding sufficient input so that it can satisfy z customers by itself and that the market has equilibrated in the manner discussed earlier. The market structure is said to be in equilibrium if at the equilibrium quantities, no stage 1 firm has any incentive to increase or decrease z. The way the market operates, there are basically two offsetting considerations involved in the decision of a stage 1 firm to produce a unit of input for itself. First, since it costs only c per unit to produce the input, the stage 1 firm will save $(p_{int} - c)$ by producing the input itself rather than buying it on the factor market at price $p_{int} > c$.[1] (In other words, if the stage 1 firm produces the input itself, then the firm assures itself of having the necessary input to make a more profitable sale, if demand should materialize.) Offsetting this saving is the potential risk the the input will be produced at cost c, but will not be used because of insufficient demand. By producing the input for itself, the stage 1 firm bears the risk of the unsold input while when the stage 1 firm relies on the factor market for the input, it is the stage 2 firms who bear this risk.

The first issue that needs investigation is whether there is any incentive for stage 1 firms to produce the input at all. So, suppose that stage 1 firms produce none of the factor input (so that $z = 0$), and rely entirely on the stage 2 markets for the input. Imagine that the stage 1 and stage 2 markets have reached equilibrium in the manner described earlier. By the conditions of market equilibrium, profits are zero in each market. Let p^o_{int} be the equilibrium price in the stage 2 market when the customer capacity z of stage 1 firms equals 0. The market structure will definitely not be stable if there is an incentive for stage 1 firms to increase z by 1 unit.

Let $\pi_i(k)$ equal the expected profits of a stage 1 firm when it holds sufficient input to satisfy k customers by itself. This $\pi_1(k)$ is calculated

---

[1] p_{int} exceeds c since p_{int} must cover not only the cost c of production, but also the cost of unsold goods. See Chapter 2.

at the prices and probabilities associated with the market equilibrium that results when $z = 0$. From the definition of equilibrium it follows that $\pi_1(0) = 0$. There will be an incentive for stage 1 firms to produce positive amounts of the factor input for themselves if $\pi_1(1) > \pi_1(0)$.

Let cost 2 = $rK + p_{int}$ and cost 1 = $rK + c$. Notice that cost 2 represents the cost of producing one unit of the final good when the factor input is purchased from a stage 2 firm at price p_{int}, while cost 1 is the resource cost when the factor input has been produced by a stage 1 firm at the price c, where c is less than p_{int}. From the definition for profits given in Chapter 2, profits for stage 1 firms can be written as

$$\pi_1(0) = [p_f - \text{cost } 2] \sum_{0}^{L} i \, pr(i) \qquad (1 - \lambda)\cdot x(p), \text{ where}$$

pr(i) = probability a firm obtains i customers, and

$$\pi_1(1) = [\sum_{i=1}^{\infty} [p_f - \text{cost } 1] + [p_f - \text{cost } 2](i - 1)(1 - \lambda)]pr(i) - pr(0)\cdot c]x(p).$$

The expression for $\pi_1(0)$ is simply the net revenue per unit times the expected number of goods that are sold. The expression for $\pi_1(1)$ is more complicated, and reflects the fact that if at least one customer appears, then the firm will be able to make a net profit on that customer of $[p_f - \text{cost } 1]$, and a net profit of $[p_f - \text{cost } 2]$ on each of the remaining customers. The term $Pr(0)\cdot c$ reflects the risk that the firm will have spent c on production of the input, yet no customers will appear to use that input. Since $\pi_1(0) = 0$, it follows that p_f = cost 2, and that

$$\pi_1(1) = [[\sum_{1}^{\infty} pr(i)][p_f - \text{cost } 1] - pr(0)\cdot c]\ x(p), \text{ or}$$

$$\pi_1(1) = [[\sum_{1}^{\infty} pr(i)][\text{cost } 2 - \text{cost } 1] - pr(0)\cdot c]\ x(p), \text{ or}$$

$$\pi_1(1) = [[\sum_1^{\infty} pr(i)][p_{int} - c] - pr(0)\cdot c]\ x(p). \qquad (2)$$

There will be an incentive for stage 1 firms to hold the input if $\pi_1(1) > \pi_1(0)$, or if $\pi_1(1) > 0$, or if

$$[1 - P(0)]p_{int} > c, \qquad (3)$$

where P(0) = the probability that at least one customer will frequent any stage 1 firm.

This inequality is intuitively plausible. If a stage 1 firm decides to stock one unit of input, its cost increases, with certainty, by c. Its expected savings from not having to go into the factor market for that one unit of input is $[1 - P(0)]p_{int}$. When savings exceed costs, the stage 1 firm will hold the factor input.

Since the demand structure is an equal share random process, we have that $[1 - P(0)] = [1 - (1 - \frac{1}{N_1})^L] \approx 1 - e^{L/N_1}$, where N_1 = number of stage 1 firms and L = total number of identical customers. Therefore, using (3), we find that stage 1 firms will integrate if

$$[1 - e^{-L/N_1}]p_{int} > c \qquad (4)$$

To determine when incentives for vertical integration exist, let us examine (4) in more detail.[1] The only variable in (4) that is endogeneous is p_{int}, which is the market clearing price in the stage 2 factor market.

[1]It should be obvious that, since the isoutility curves must only satisfy an upper and lower Lipshitz condition, then for any L, N_1, N_2 and c, there is always some set of preferences that yield a market equilibrium (i.e. tangency between isoutility and zero profit curves) with a p_{int} such that (4) holds.

This price depends on the number of stage 2 firms, N_2, the cost, c, of producing the input, and the preferences and the number, L, of consumers. It does not depend at all on the number of stage 1 firms, N_1. (Recall that the price of the factor, p_{int}, exceeds the resource cost, c, since in equilibrium price has to cover not only production costs, but also the cost of unsold goods.)

It is not possible to determine whether (4) is more likely to hold as the number of customers, L, increases without explicitly specifying the tastes of consumers. (This is not as true of an increase in the customer per firm ratio, L/N_1.) As L increases, $1 - e^{-L/N_1}$ approaches 1, so that incentives to integrate increase. However, from Chapter 2, we know that as L increases the equilibrium, p_{int} approaches c, so that disincentives to integrate are increased. Whether (4) is more likely to hold as L increases depends on how fast p_{int} approaches c, which depends on the specific slopes of the isoutility curves. However, it is clear from (4), that for fixed L, (4) is more likely to hold the smaller is the number, N_1, of stage 1 firms. (Of course, by the assumptions of the model $N_1 > N_2$). Equivalently, for a fixed number of customers, L, (4) is more likely to hold the larger is the customer per store ratio L/N_1.

This last result emphasizes the importance of examining the uncertainty in the market clearing process. In Chapter 2, it was established for large customer per firm ratios, that, at least percentagewise, the market equilibrium under uncertainty approached that under certainty. In such a case, the reader might have thought that the deterministic analysis which ignored the uncertainty in the market would suffice, and the more complicated analysis which explicitly considered the uncertainty was unnecessary. What we see here though is that it is precisely the case of having a large customer per firm ratio in stage 1 markets that can lead to strong incentives for vertical integration. A deterministic analysis of this market structure would have been unable to find any incentives or disincentives for vertical integration to occur.

The implications of (4) are disturbing. It might initially have seemed that having a small number of firms in either the stage 1 or stage 2 markets is socially desirable, since then the "riskiness" of the demand structure facing individual firms (which in the models under study is inversely related to the customer per firm ratio) is reduced. However, as long as the number of stage 1 firms exceeds the number of stage 2 firms[1], Theorem 1 tells us that it is socially undesirable to allow vertical integration. For such cases, (4) tells us that the smaller the number of firms N_1 in the stage 1 markets, the stronger is the incentive for the socially undesirable vertical integration to occur. Therefore, the initial reasoning about the desirability of having a small N_1 needs to be changed. As long as the number of stage 1 firms exceeds the number of stage 2 firms, it is socially desirable to have a _large_ number of stage 1 firms, so that the customer per store ratio in stage 1 markets, L/N_1 is _not_ large. Then (4) is likely not to hold and no incentives for the socially undesirable vertical integration will exist. What does remain true about the initial intuition is that having a small number of stage 2 firms is always socially desirable.

The incentives for vertical integration come about because the stage 1 firms base their decisions to integrate on the marginal, not average, probability of using an additional input. The way the markets operate, the price

[1]We discuss later the situation where the number of stage 2 firms exceeds the number of stage 1 firms.

of the factor in the stage 2 market reflects not only the cost c of producing the input, but also the average probability of not being able to sell that input. When a stage 1 firm is deciding whether to hold one unit of the input itself, is is not concerned with the average probability of being unable to use any unit of input. Rather, since the stage 1 firm will use its input holdings first, the stage 1 firm is concerned with the probability of being able to use that first unit of input. For even low to moderate values of the customer per firm ratio in stage 1 markets, L/N_1, (e.g. 15-30), this probability is practically 1, (i.e. each stage 1 firm is virtually assured of being able to use up its one unit of input), so that (4) will almost certainly hold since the price of the stage 2 factor, p_{int}, exceeds c. It is precisely because stage 1 firms can use their own input to satisfy their "high probability" demand and use the stage 2 market to satisfy their "low probability" demand that incentives for vertical integration occur. The disturbing conclusion of this analysis is that it is quite likely that there will exist strong private incentives for socially undesirable vertical integration to occur.

5.7 Equilibrium Market Structure

If at the equilibrium associated with stage 1 firms providing customer capacity z, there is no incentive for stage 1 firms to increase or decrease z, then the market structure is said to be in equilibrium. If (4) does hold, then the equilibrium must involve some vertical integration.

For example, suppose the equilibrium market structure involves partial vertical integration. Each stage 1 firm produces sufficient input to satisfy z customers by itself and p_{int} and $1 - \lambda_2$ are the resulting equilibrium price and probability of satisfaction in the stage 2 market. The stage 1 firms will have no incentive to alter z if there does not exist a customer capacity, y, and price, p_f, such that at the equilibrium quantities, $1 - \lambda_2$, p_{int} the following is true: (a) at that y and p_f, the firm is able to offer consumers a higher level of utility than they are enjoying in the current market equilibrium; and (b) the stage 1 firm could earn non-negative profits. Let $\bar{u}$ stand for the level of utility achievable in the equilibrium when stage 1 firms each provide a customer capacity z, and π_1 stand for stage 1 profits. Then for easy reference, we can write that for equilibrium involving partial vertical integration stage 1 firms have no incentive to alter their customer capacity, z, if there does not exist a y and p_f such that

$$U(1 - \lambda_2 \cdot \lambda_1(y), p_f) > \bar{u} \text{ and } \pi_1(p_f, y/p_{int}, 1 - \lambda_2) \geq 0 \qquad (5)$$

For equilibrium market structures involving partial vertical integration, the market structure is said to be in equilibrium, with each stage 1 firm providing customer capacity z for itself, when (5) holds. As will soon be seen, the incentives for vertical integration neet not depend monotically on z, because of the influence of z on the equilibrium price p_{int} and probability of satisfaction, $1 - \lambda_2$, in the stage 2 market. Therefore, it might be possible to have several different equilibrium market structures that satisfy (5).

Suppose that (4) holds, so that each stage 1 firm finds it profitable to produce enough of the factor input for itself so as to be able to satisfy the demands of one customer. Will the stage 1 firms find it profitable to continue to expand their own production of the factor input? The answer to the above question depends on the new market equilibrium reached as a result of the stage 1 production of the factor. The Lemma below establishes one form that the incentive conditions can take.

Lemma 1: A stage 1 firm will continue to expand its customer capacity from z to z + 1, if

$$-c + \left[p_{int} + \lambda_2[p_f - \text{cost } 2]\right][1 - P(z)] > 0 \qquad (6)$$

where P(z) = cumulative probability that fewer than z + 1 customers will visit a stage 1 firm, and all of the other previously defined variables are at their equilibrium values that result when each stage 1 firm produces sufficient input to satisfy z customers by itself. For notational convenience, the expression on the left hand side of (6) will be referred to as Inc(z) (for incentive at z), so that (6) can be written as

$$\text{Inc}(z) > 0.$$

Proof: Appears in the appendix to this chapter, Appendix C.

To interpret Inc(z), rewrite (6) as

$$\text{Inc}(z) = -c \cdot P(z) + \left[[p_f - \text{cost } 1] - [p_f - \text{cost } 2](1 - \lambda_2)\right](1 - P(z)) \qquad (7)$$

The first terms reflects the expected cost of being unable to use the additional input. The second term reflects the expected gain of holding the additional input. This expected gain consists of the difference of two terms. The first is the profit that the firm will definitely make if more than z customers show up and it holds the additional input. Subtracted from this term is the expected net revenue that it would have made if it did not have the additional input and had to purchase the input from the factor market.

Lemma 1 showed that if the expected gains from producing an extra unit of factor input exceed the cost so that Inc(z) is positive, a stage 1 firm will desire to integrate further. If Inc(0) $\geq$ 0, so that there is an incentive for vertical integration, does the incentive for vertical integration persist until the factor markets are forced otu of existence? In other words, do we have the situation, common in models of vertical integration, that if it is ever profitable to integrate, then complete vertical integration will occur with the factor markets disappearing and every stage 1 firm relying on itself for production of the factor input? To answer this question, we must examine the possible types of equilibrium market structure that can occur.

There are three types of equilibrium market structures that are possible in the model under discussion. The first is the socially preferred case involving no vertical integration. For this case to occur (4) must not be satisfied. The second type is the incomplete vertical integration case. This case occurs when (4) is satisfied, and at some z, (5) is satisfied. Moreover, referring back to the way in which equilibrium is derived, a requirement for incomplete integration is that the utility level of the derived isoutility curve, associated with the equilibrium point of tangency in $(1 - \lambda_2, p_{int})$ space is higher than the utility level that occurs when the stage 1 firms do not use the stage 2 market and rely entirely on themselves for production of the input. The final type of market structure equilibrium is the complete integration one, where the stage 2 markets disappear. This market structure occurs when the incentives for vertical integration cause stage 1 firms to increase their customer capacity to a point where the utility level that the stage 1 firm can provide by using the stage 2 market is lower than that which would result if the stage 1 firm produced its own input and relied entirely on itself for its input supply.

If Inc(0) > 0, then vertical integration must occur but, a priori, without specifying in detail consumer preferences, it is impossible to tell whether the second or third type of market structure equilibrium will result.[1] The difficulty arises because the incentives for increased vertical integration depend on the equilibrium prices and probability of satisfaction, both of which depend on the specific shape of the consumers' isoutility curves. For example, as z increases (i.e. as stage 1 firms continue to vertically integrate) the incentives for vertical integration, as expressed in (6), may either increase or decrease depending on the particular preferences of customers. The probability of obtaining at least z customers, [1 - P(z)], will definitely decrease as z increases. However, the amount by which the equilibrium prices, p_f, p_{int}, and probability of

[1]As the reader is probably aware, the equilibrium conditions as expressed in (5) are analytically very complicated to deal with. This makes the determination of the equilibrium market structure a difficult question. However, there are several comments that can be made about equilibrium market structure. It should be obvious that a market structure involving no vertical integration is possible. Simply choose L, N_1, and N_2 so that L/N_2 is large (so that $p_{int} \approx c$) and choose N_1 very large so that L/N_1 is small (so that the risk to stage 1 firm of being unable to use their own input is huge). Moreover, as mentioned in the footnote to (4), it is also clear that it is possible for incentives for vertical integration to exist. When such incentives exist, it is impossible, using (5) to determine whether the integration will be partial or complete. If $N_1 \approx N_2$, so that there is little difference in the risk absorbing capabilities between stage 1 and stage 2 firms, the integration is likely to be complete. On the other hand, if $N_1 >> N_2$, we expect the integration to be only partial. The reader should recall that there are virtually no restrictions on the shape of consumer preferences. With such freedom, it would be very surprising if it were not possible for either type of vertically integrated equilibrium to exist within the model. Both types of vertical integration seem to exist in the real world for reasons regarding assurance of supply that are precisely the issues that this model attempts to address. In the subsequent analysis, we will assume that is possible for both types of vertical integration to occur in the model.

satisfaction in the stage 2 markets, $1 - \lambda_2$, will change in response to increases by stage 1 firms in the customer capacity, z, will depend on the particular shape of consumer preferences. Furthermore, whether the utility level achievable when stage 1 firms use the stage 2 markets exceeds that achievable when stage 1 firms rely entirely on themselves for production of the factor input will depend once again on consumer preferences.

Before dealing directly with the differences in equilibrium between the first and the last two types of market structures, let us examine the second and third types in a little more detail. Suppose the equilibrium market structure is of the second type involving incomplete vertical integration, with each stage 1 firm producing enough factor inputs for itself to satisfy z customers. Then, the stage 2 market acts like an insurance market for supplying the factor input to the stage 1 market. To see this last point, notice that whenever a stage 1 firm makes a sale of a final product, it makes a higher per unit profit when it is able to use its own input (produced at cost c) in the manufacture of the final good rather than when it uses an input purchased on the stage 2 market at a price p_{int} (which exceeds c). A stage 1 firm continues to enter the stage 2 market simply because it needs to satisfy its customers. It is cheaper for a stage 1 firm to satisfy its customer through use of the high price stage 2 market, rather than produce extra input for itself and bear the risk that the unit of input will go unsold.

Based on the foregoing discussion, we observe that there is one case where the strong private incentive for vertical integration is socially desirable. Suppose that the number of stage 2 firms exceeds the number of stage 1 firms, so that it is more efficient to have the stage 1 firms absorb the risk of producing the factor input. For this case, it is immediate from the preceding analysis that the private incentives for vertical integration will bring about the equilibrium market structure of complete vertical integration with the stage 2 firms disappearing. Since, in this case, stage 1 firms are more efficient absorbers of risk, it follows that consumers are better off in the market structure involving complete vertical integration than in any other market structure. In other words, consumers prefer when the stage 1 firms rely entirely on themselves for production of the input, and bear the entire risk of having unsold input. However, it seems more natural to expect that the factor input stage 2 firms will be better absorbers of risk than the final product stage 1 firms. For the competitive markets under study, we usually would expect that any firm producing factor inputs is producing a large enough amount of inputs to satisfy, by itself, several of its potential final product customers. (In fact, this last notion underlies the "foreclosure" theory, which has often been used as an argument in antitrust suits against vertical integration. In the "foreclosure" theory, backward vertical integration is deemed undesirable because the vertically integrating firm usurps more production capacity than it actually needs, and thereby reduces the input supply opportunities for competing final product firms.)[1] Unless otherwise noted, we will continue to conduct the analysis on the assumption that the stage 2 firms are more efficient absorbers of risk than the stage 1 firms. As shown in Section 5.5 for this case, the private incentive to vertically integrate is socially undesirable. Section 5.14 will deal more explicitly with the issues of the relative risk absorbing efficiency of stage 1 and stage 2 firms.

---

[1]For more on the foreclosure thoery, see S. Peltzman and J. Weston, ed., Government Policy Toward Mergers, Goodyear Publishing Co., 1968.

5.8 The Consequences of Vertical Integration on Market Equilibrium

In this section, we examine how the market equilibrium is affected when the equilibrium market structure involves some vertical integration. We compare the market equilibrum with vertical integration to the market equilibrium when vertical integration is not allowed. From Theorem 1, we already know that, compared to the expected level of utility achieved in the competitive equilibrium with no vertical integration, the expected level of utility is lower in any market structure involving vertical integration. Since vertical integration lowers the expected level of utility, it is possible that the level of utility could be driven so low that consumers would prefer not to enter the stage 1 market. Thus, one consequence of vertical integration can be to drive markets out of existence. Having mentioned this possibility, we shall concentrate in the subsequent analysis on the effects of vertical integration when the markets under study remain in existence.

The questions we ask are whether the price, p_f, of the final product the price, p_{int}, of the factor sold in the stage 2 market, and the probability of satisfaction, $1 - \lambda$, are higher or lower in a vertically integrated market structure than in a market structure in which vertical integration is not allowed. This section is divided into two parts, the first discusses the rationale for the assumptions that are made and presents a heuristic discussion of the main results and the reasoning underlying them. The second subsection is more technical and presents the assumptions behind the proofs of the main results in greater detail.

5.8.1 Assumptions and Results

From Chapter 2, we know that market equilibrium depends on both the firms' zero profit curve and the consumers' isoutility curves. If we make a small change in the shape of the isoutility curves, we expect that a small change in the market equilibrium will occur. We therefore make an assumption that essentially says that small changes in preferences will not lead to huge jumps in the market equilibrium quantities. This assumption enables us to use marginal incentives to predict how the market equilibrium will change in response to a change in market structure.

The next assumption we make is analogous to the assumptions of "normal" goods in economic theory. We assume that if the customer per firm ratio increases so that the consumers are offered a better "menu" of price-probability of satisfaction combinations,[1] then the consumer will prefer a combination in which he is made better off in both dimensions. In other words, the consumer will choose a combination with a lower price and higher probability of satisfaction.

This assumption appears very reasonable since in Chapter 2, we proved that as the customer per firm ratio increases, the market equilibrium price approaches its minimum possible value of c, and the probability of satisfaction approaches its maximum value of one. Hence the consumer must be made better off in terms of both the price and probability of satisfaction for sufficiently large increases in the customer per firm ratio. To see the analogy of the above assumption to consumer theory, note that we usually

[1] Recall that as the customer per store ratio increases, the zero profit curve, used to define equilibrium, shifts up so that consumers are faced with a better set of price-probability of satisfaction choices.

assume that a consumer in a two good world would increase his consumption of both goods in response to an increase in income (i.e. as his "menu" between the two goods improves).

Suppose that the equilibrium market structure involves complete vertical integration. Using the above assumptions, it is possible to establish how the market equilibrium with complete vertical integration differs from the market equilibrium when vertical integration is not allowed. Since the stage 2 factor market disappears with complete vertical integration, we need only compare the price of the final good, p_f, and the probability of satisfaction $1 - \lambda$. The main result is that the market equilibrium with complete vertical integration involves a higher price and lower probability of satisfaction than does the market equilibrium with no vertical integration. Stage 1 firms are less efficient absorbers of risk than stage 2 firms in the sense that stage 1 firms with complete vertical integration have to spend more resources on production of inputs than do stage 2 firms, with no vertical integration, to satisfy any given fraction of the population. The final price to the consumer has to rise to cover this increased cost of operation in the case of complete vertical integration. Moreover, because stage 1 firms cannot satisfy customers as efficiently as stage 2 firms, the equilibrium probability of shortage in the case of complete vertical integration rises from its value in the market equilibrium when vertical integration is not allowed. From these two results, it follows that the total amount of the output that is purchased is lower in the case involving complete vertical integration than in the case involving no vertical integration.

Suppose that the equilibrium market structure involves partial vertical integration, let us compare the market equilibrium with partial vertical integration to the market equilibrium when no vertical integration is allowed. The main result is that the price of the input purchased in the stage 2 market is higher in the case involving partial vertical integration than in the market equilibrium when no vertical integration is allowed. This result follows from the fact that in the case of partial vertical integration, the stage 2 markets become "riskier"[1] and the stage 2 firms become less efficient absorbers of risk than they were in the case of no vertical integration. This inefficiency results in increased costs to stage 2 firms. To cover their increased costs, the stage 2 firms have to raise their prices to the stage 1 firms. Surprisingly, it does not appear possible to prove that the stage 1 firms pass this increased cost along to the consumer in terms of higher prices for the final good. It seems possible, though I suspect unlikely, that with partial integration, the price of the stage 2 input could rise, but the price of the final good could fall. In this case, we know from Theorem 1 that the probability of satisfaction would have to fall sufficiently so that consumers are worse off in the case of partial vertical integration than in the case of no vertical integration.

The proof of the result that the price of the stage 2 input will increase with partial vertical integration is considerably more difficult than the proofs of the previous results comparing complete vertical

---

[1]Recall that "risk absorbing" efficiency is inversely related to the customer per firm ratio. When stage 1 firms produce some of their own input, the stage 2 firms see less than L customers, so that the customer per firm ratio of stage 2 firms falls from its value in the case of no vertical integration.

integration to no vertical integration. The proof becomes difficult because it is necessary to trace the effect of partial integration on the way the derived isoutility curves shift.

(As described in Section 5.4, it is possible to construct "derived" isoutility curves that express consumer preferences for the price, p_{int}, of the stage 2 factor, and the probability of satisfaction, $1 - \lambda_2$, in the stage 2 market. These derived isoutility curves reflect the tradeoffs in the factor market that translate into tradeoffs in $(1 - \lambda, p_f)$ space that leave the consumer indifferent. Recall that the tangency between the stage 2 zero profit curve and the highest derived isoutility curve establishes equilibrium in the stage 2 market.)

To prove the result about the input price rising, it is necessary to assume that the derived isoutility curves shift in a particular way when the market structure becomes partially integrated. Basically, the assumption states that the probability of satisfaction per unit of customer capacity provided by partially integrated stage 1 firms, is less than that provided by the more efficient risk absorbing stage 2 firms in the nonintegrated market. This assumption is most likely to hold when the number of stage 1 firms greatly exceeds the number of stage 2 firms. This latter situation is the case of most interest to a policy maker, since the costs of vertical integration become more severe, the greater the differential in the ability of stage 1 and stage 2 firms to absorb risk.

The final assumption needed to prove the result about the input price rising deals with the concavity of consumer preferences and the slopes of the derived isoutility curves. Recall from Chapter 2 that concavity is a plausible (though certainly not compelling) assumption to impose on consumer preferences. Although it can be argued that the previous assumptions of this section, especially those used in the comparison between complete and no vertical integration, are very weak (i.e. extremely plausible), the same cannot be said of this final assumption.

In summary then, any market structure involving vertical integration provides a lower level of utility to consumers than the market structure involving no vertical integration. Any vertical integration causes an inefficiency in the ability of firms to absorb risk, and usually will result in higher prices in the input market. When the equilibrium market structure involves complete vertical integration, we expect that both the probability of shortage and the price of the final good will rise from their equilibrium values in the case when integration is not allowed.

5.8.2 Comparison of Market Equilibrium With and Without Vertical Integration*

From the discussion in Section 5.4 on how the markets under analysis operate, it is obvious that to obtain any answers to the questions of how market equilibrium differs with and without vertical integration, it will be necessary to make some assumptions about the shape of the isoutility curves and about some general properties of market equilibrium.

Assumption 1: Consider the isoutility curves which reflect consumer preferences between the probability, $1 - \lambda$, of obtaining the good, and the price, p_f, of the final good. In $(1 - \lambda, p_f)$ space, for any fixed p_f, the slope of the isoutility curves falls as the level of utility falls.

*This section may be omitted by the reader not interested in the technical details of the assumptions and proofs that underlie the results described in the previous subsection. The first part ofthis subsection deals with the comparison between complete and no vertical integration. The second part deals with the comparison between partial and no vertical integration. The second part is more technical (and tedious) than the first.

Actually, the above statement is called an assumption to avoid any arguments with someone who may not believe in Von Neumann utility functions. If one is willing to accept the existence of such a utility function, then Assumption 1 is actually a theorem, whose proof appears below.

Proof: Let $u(x_1,x_2)$ be the Von Neumann utility function, where x_1 is the good subject to shortages and x_2 is a good always available at price 1. Then, as explained in Chapter 2, expected utility can be written as

$$U(1 - \lambda,p) = (1 - \lambda)V(p,Y) + \lambda \cdot u(0,Y) \quad \text{where}$$

$1 - \lambda$ = probability of obtaining good 1

Y = endowment income

p = price of good 1

$V(p,Y)$ = indirect utility function

For any given level of expected utility $U = \bar{u}$, $(u(0,Y) \leq \bar{u} \leq V(p,Y))$, the isoutility curves are given by

$$1 - \lambda = \frac{\bar{u} - u(0,Y)}{V(p,Y) - u(0,Y)} \tag{8}$$

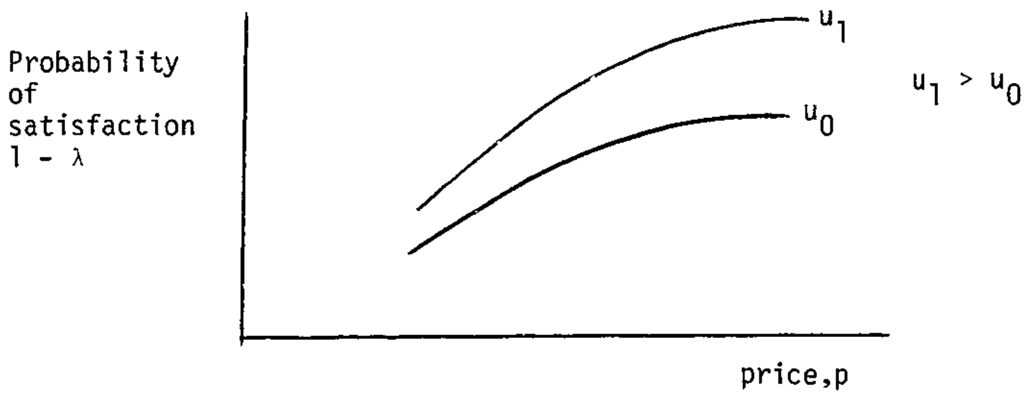
Differentiating (8), one obtains

$$\frac{d(1 - \lambda)}{dp} = (\bar{u} - u(0,Y)) \frac{1}{[V(p,Y) - u(0,Y)]^2} \frac{(-1)}{} V_p(p,Y) \tag{9}$$

where the subscript represents partial differentiation.

From (9), it is evident that for fixed p, as $1 - \lambda$ (or equivalently $\bar{u}$) falls, the slope of the isoutility curve also falls. Diagramatically, the situation is depicted below.

Figure 5-5 The Slope of the Isoutility Curves



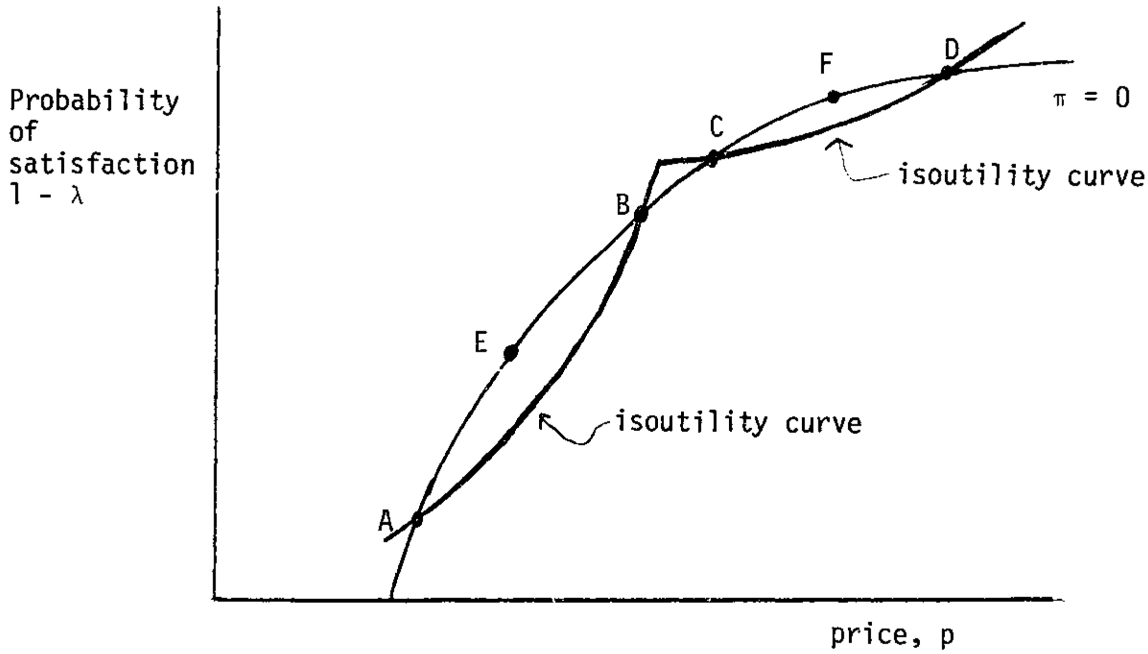
The next assumption deals with the relationship between the competitive market equilibrium and the slopes of the zero profit and isoutility curves at points other than the competitive equilibrium.

Assumption 2: Consider competitive equilibrium as defined in Chapter 2, as the tangency between the zero profit (i.e. $\pi = 0$) curve of firms and the highest isoutility curve. Suppose at some point A on the $\pi = 0$ curve that the slope of the isoutility curve is flatter than that of the $\pi = 0$ curve. Then, the price associated with the equilibrium point exceeds the price associated with point A.

The purpose of this assumption is to rule out situations where marginal incentives do not lead to the correct global incentives. More specifically, when only marginal changes are possible, it is possible to construct examples where the market could get "stuck" at the wrong point. For example, consider the diagram in Figure 5-6.

Although point E in Figure 5-6 may lead to the highest isoutility level, if firms start operating at C or D, then they will always face marginal incentives to move toward point F, and not toward point E. Since in most economic analysis only marginal incentives can be examined, it is necessary to create a link between marginal and global incentives in order to reach conclusions about market operation. Assumption 2 does just that, and rules

Figure 5-6 Marginal and Global Incentives



out cases where nonmarginal changes would lead to a different equilibrium than would marginal changes. Basically, Assumption 2 captures the idea that marginal changes in either the $\pi = 0$ curve or the shape of the isoutility curve will lead to marginal, not discrete, changes in the market equilibrium. Therefore, it is possible to predict what a new equilibrium will look like in response to altered conditions by examining the marginal incentives created under the new circumstances.

There are several alternative ways of establishing the same implications as Assumption 2. For example, convexity of isoutility contours, or the requirement that isoutility curves be of such a shape that they intersect a $\pi = 0$ curve at most twice would achieve the same objective as Assumption 2.

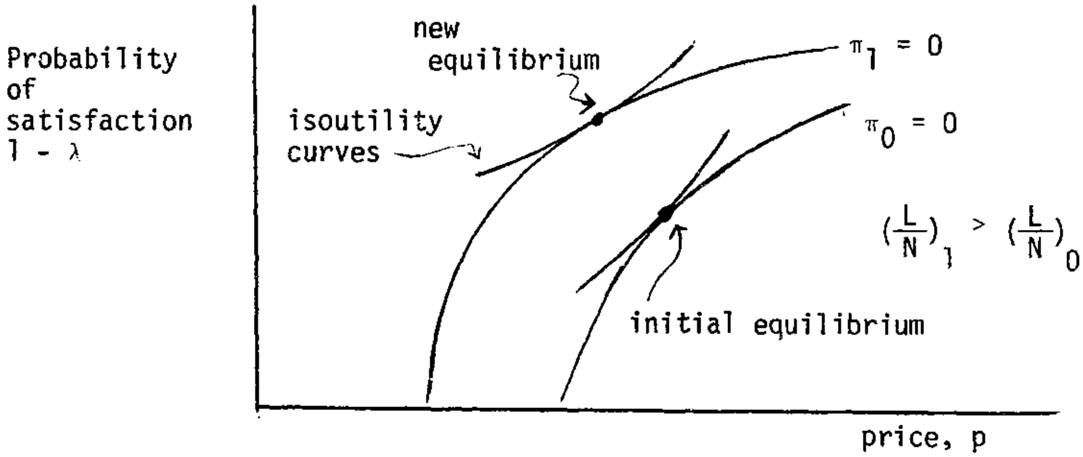
The next assumption deals with how the market equilibrium behaves as the customer per firm ratio increases and is analogous in consumer theory to the assumption of a "normal" good. In Chapter 2, equilibrium was defined as the tangency of the highest isoutility curve with the zero profit (i.e. $\pi = 0$) curve. Suppose that the customer per firm ratio, L/N, increases, so that the $\pi = 0$ curve shifts up. In the new equilibirum the consumer is better off than before. The question arises as to whether the probability of satisfaction $1 - \lambda$ rises and/or the price, p, falls in the new equilibrium.

The question asked here is analogous to the question of whether a consumer in a two good economy will increase his consumption of both goods in response to an increase in income. In consumer theory, the normal case is to assume that the consumer increases his consumption of both goods. Similarly, in this situation, it would seem that the normal case would involve the consumer being made better off in both the $1 - \lambda$ and p dimensions. The plausibility of such an assumption is increased, when one realizes that the theorems of Chapter 2 tell us that for sufficiently large increases in the customer per firm ratio, the consumer _must_ be made better off in both the $1 - \lambda$ and p dimensions. For future reference, we formalize the above reasoning as Assumption 3.

Assumption 3: Consider the competitive market equilibrium as defined in Chapter 2. Let the customer per firm ratio increase, so that the zero profit ($\pi = 0$) curve, used to define equilibrium, shifts up. In the new equilibrium, the price p, is lower and the probability of satisfaction, $1 - \lambda$, is higher than in the initial equilibrium. Assumption 3 is illustrated below.

Given Assumptions 1 through 3, we can now investigate the differences in the equilibrium quantities between the nonvertically integrated and vertically integrated cases. As explained in Section 5.7, there are two

Figure 5-7 Equilibrium and the Customer Per Firm Ratio



types of equilibrium market structures involving vertical integration. We will first examine the case when the equilibrium market structure involves complete vertical integration, in which the stage 2 market disappears. We compare the market equilibirum with complete vertical integration, to the one that results when no vertical integration is allowed.

When there is no vertical integration, equilibrium is determined as the tangency between the zero profit curve for stage 2 firms (i.e. $\pi_2 = 0$) and the derived isoutility curves in $(1 - \lambda, p_{int})$ space, where $rK + p_{int} = p_f$ represents the zero profit condition for stage 1 firms.[1] The π_2 curve refers to the zero profit curve in the stage 2 market and is drawn on the assumption that the number of stage 2 firms is N_2 and the number of customers is L. In the case of complete vertical integration, the stage 2 firms disappear, so the stage 1 firms must provide the input themselves. For this case, equilibrium is determined by the tangency between the isoutility curves and the $\pi_1 = 0$ curve in $(1 - \lambda, p_{int})$ space, once again. The $\pi_1 = 0$

[1]Notice that when there is no vertical integration, the derived isoutility curves in $(1 - \lambda, p_{int})$ space are identical to the isoutility curves in $(1 - \lambda, p_f)$ space except for a horizontal translation by rK units. (See Section 5.4 for the definition of derived isoutility curves.)

curve is the zero profit curve in the stage 2[2] market and is drawn on the premise that there are N_1 firms and L customers. Since $N_1 > N_2$, the $\pi_1 = 0$ curve lies entirely below the $\pi_2 = 0$ curve.

Theorem 2: As compared to the nonvertically integrated equilibrium, the equilibrium involving complete vertical integration has the following properties:

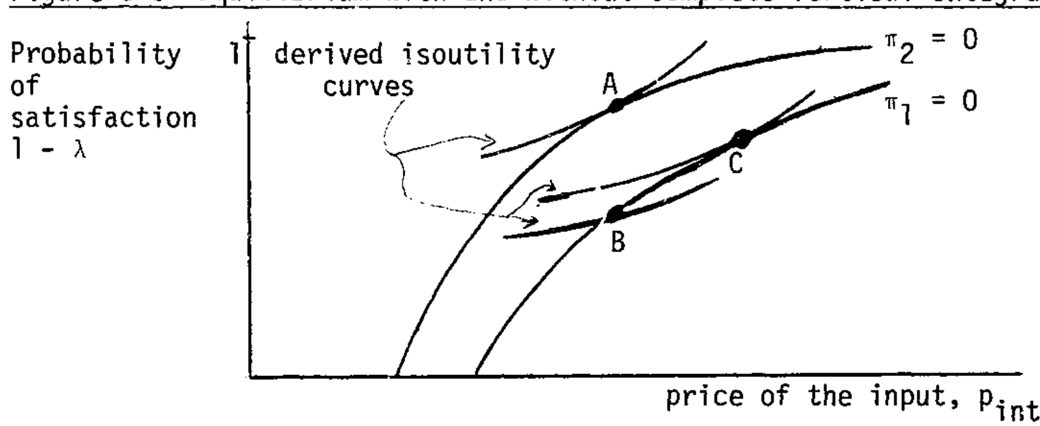
(a) under Assumptions 1 and 2, the price of the stage 2 input, p_{int}, and the price of the final good, p_f, are higher;

(b) under Assumptions 1 - 3, the probability of satisfaction, $1 - \lambda$, is lower; and

(c) under Assumptions 1 - 3, the total amount sold of the final good is less.

Proof: Since $L/N_1 < L/N_2$, the $\pi_1 = 0$ curve lies completely below the $\pi_2 = 0$ curve. Let equilibrium in the nonvertical integration case occur at some point A, depicted below.

Figure 5-8 Equilibrium With and Without Complete Vertical Integration



[1] Even though there is no stage 2 market in the case of complete vertical integration, the stage 1 firms effectively act as stage 2 firms when they produce the input. That is why even in the case of complete vertical integration market, equilibrium can be viewed as the tangency between the derived isoutility curve and a zero profit curve in the stage 2 market.

By the theorems in Appendix A, we know that the slope at point B along the $\pi_1 = 0$ curve is steeper than that at point A along the $\pi_2 = 0$ curve. But from Assumption 1, the slope of the isoutility curve at point B is less than that at point A. Hence, if there was a tangency between the $\pi_2 = 0$ and isoutility curve at point A, then at point B the slope of the isoutility curve is less than the slope of the $\pi_1 = 0$ curve. This implies, by Assumption 2, that the equilibrium in the case of complete vertical integration lies to the right of point B in the above diagram at some point C, or that in the equilibrium with complete vertical integration at point C, p_{int} at point C exceeds p_{int} at point A. Since the zero profit condition in stage 1 markets requires that $p_f = rK + p_{int}$, part (a) of the theorem follows immediately.

By using the above reasoning, and applying Assumption 3, part (b) of the theorem follows immediately.

To prove part (c), let a subscript 1 and 2 stand for the equilibrium quantities in the case of complete vertical integration and no vertical integration, respectively. Then, it follows from parts (a) and (b) of the theorem that

$$p_{f_1} > p_{f_2} \Rightarrow x(p_{f_1}) < x(p_{f_2})^1,$$

and $$(1 - \lambda)_1 < (1 - \lambda)_2,$$

so that $$L(1 - \lambda)_1 x(p_{f_2}) < (1 - \lambda)_2 x(p_{f_2})L$$

or $$(\text{total amount sold})_1 < (\text{total amount sold})_2$$ Q.E.D.

[1]This assumes that the demand curve is downward sloping -- i.e. $x'(p_f) < 0$.

Notice that Theorem 2 does not say how the total amount of input that is produced each period compares between the two different market structures. Let a subscript 1 and 2 again denote the equilibrium quantities in the case of complete integration and no integration respectively, and let s stand for the number of customers each firm can service. It is possible for the total amount of the factor produced $x(p_{f_1}) \cdot N_1 \cdot s_1$ to be either greater or less than $x(p_{f_2}) \cdot N_2 \cdot s_2$. The reason for this is that although Theorem 2 tells us that the probability of satisfaction $(1 - \lambda)_1$ is less than $(1 - \lambda)_2$, it is possible for total customer capacity $N_1 \cdot s_1$ to be either greater or less than total customer capacity $N_2 \cdot s_2$. Either relationship is consistent with $(1 - \lambda)_1 < (1 - \lambda)_2$. (Initially, one might think that $(1 - \lambda)_1 < (1 - \lambda)_2$ implies that $N_1 \cdot s_1 < N_2 \cdot s_2$. This need not be the case since stage 1 firms are less efficient absorbers of risk than stage 2 firms. Hence, in the complete integration case, stage 1 firms could have a larger customer capacity, $N_1 s_1$, but still have a lower probability of satisfying a customer, $(1 - \lambda)_1$, than the stage 2 firms in the no integration case.)

Let us examine the market equilibrium when the equilibrium market structure involves partial vertical integration. We compare the market equilibrium with partial vertical integration to the one that results when vertical integration is not allowed. This comparison will be more difficult to make than the previous one and will depend on the complicated shifting that occurs in the shape of the derived isoutility curves in response to vertical integration. (See Section 5.4 for a discussion of the derived isoutility curves.) To understand in more detail how market equilibrium with partial integration compares to that with no vertical integration, it is necessary to establish some properties for derived isoutility curves.

Lemma 2: Consider the point in $(1 - \lambda_2, p_{int})$ space that maps into some $(1 - \lambda^*, p_f^*)$[1], as discussed in Section 5.4, when stage 1 firms each produce enough input to satisfy z customers by themselves. Call the slope of the derived isoutility curve at this point in $(1 - \lambda_2, p_{int})$ space m_1. Now, consider the derived isoutility curves when stage 1 firms rely entirely on stage 2 firms for their supplies of the factor input. Consider the point in $(1 - \lambda_2, p_{int})$ space that maps once again into $(1 - \lambda^*, p_f^*)$. Call the slope of the derived isoutility curve at this point m_2. Then, m_2 is less than m_1.

Proof: The isoutility curves in $(1 - \lambda, p_f)$ space are given by $U(1 - \lambda, p_f) = \overline{u}$. Let $m_* = (-1)U_2/U_1$ = the slope of the isoutility curve at $(1 - \lambda^*, p_f^*)$, where a subscript denotes partial differentiation.

With no vertical integration, the derived isoutility curves are drawn in $(1 - \lambda, p_{int})$ space and p_f and p_{int} are related by the stage 1 zero profit condition and $rK + p_{int} = p_f$. Hence, with no vertical integration, we have that $m_2 = m_*$ at the point in $(1 - \lambda, p_{int})$ space that maps into $(1 - \lambda^*, p_f^*)$ (i.e. at the point $(1 - \lambda^*, p_f^* - rK)$) since

$$\frac{dp_f}{dp_{int}} = \frac{d1 - \lambda}{d1 - \lambda_2} = 1.$$

Now consider the case involving partial vertical integration. Suppose that the stage 1 firms can satisfy their customers with probability $1 - \lambda_1$ by using their own input holdings. Then, a point on the derived isoutility

[1]To refresh the reader's mind, $1 - \lambda_2$ is the probability of satisfaction in stage 2 markets, p_{int} is the price of the input sold in the stage 2 market, $1 - \lambda$ is the probability that a stage 1 firm can satisfy a customer, and p_f is the price of the final product sold in the stage 1 market.

curves in $(1 - \lambda_2, p_{int})$ space is related to its image in $(1 - \lambda, p_f)$ space by the following questions:

$$1 - \lambda = 1 - \lambda_1 \cdot \lambda_2, \text{ and} \tag{10}$$

$$0 = (1 - \lambda)p_f \frac{L}{N_1} - cz - (1 - \lambda)\frac{L}{N_1} r \cdot K - \lambda_1(1 - \lambda_2)\frac{L}{N_1} p_{int} \tag{11}$$

where both relations were derived previously in Section 5.4.

For the case involving partial vertical integration, the slope of the dervied isoutility curve, when evaluated at the point in $(1 - \lambda_2, p_{int})$ space that maps into $(1 - \lambda^*, p_f^*)$ equals

$$m_2 = \frac{d\ (1 - \lambda_2)}{dp} = \frac{U_2}{U_1}\ (-1)\ \frac{d\ p_f}{d\ p_{int}}\ \frac{1}{\frac{d\ (1 - \lambda)}{d(1 - \lambda_2)}}, \text{ or}$$

$$m_2 = m_* \frac{1(1 - \lambda_2)}{1 - \lambda} \frac{\frac{L}{N_1}}{\frac{L}{N_1}} \frac{1}{\lambda_1}, \text{ or}$$

$$m_2 = m_1 \frac{1 - \lambda_2}{1 - \lambda}, \text{ or}$$

$$m_2 < m_1,$$

since $1 - \lambda_2$ is less than $1 - \lambda$. Q.E.D.

Using Lemma 2 and (10) and (11), let us try to see how a point and its associated slope in derived isoutility space shift when vertical integration occurs. Choose any $(1 - \lambda^\circ, p_f^\circ)$ combination, and consider the corresponding point in derived isoutility space (i.e. in $(1 - \lambda_2, p_{int})$ space). First, consider the case of no vertical integration (i.e. $\lambda_1 = 1$, $z = 0$). From (10) and (11), we find that the point A° in derived isoutility space that corresponds to $(1 - \lambda^\circ, p_f^\circ)$ is given by

$$p^{\circ}_{int} = p^{\circ}_{f} - rK \tag{12}$$

and
$$1 - \lambda^{\circ}_{2} = 1 - \lambda^{\circ}. \tag{13}$$

Now, suppose that the equilibrium market structure involves partial vertical integration and that each stage 1 firm provides customer capacity z for itself in this equilibrium, and can satisfy $1 - \lambda_1$ of its customers by itself.[1] Again, using (10) and (11), we find that the point A** in derived isoutility space that now corresponds to $(1 - \lambda^{\circ}, p^{\circ}_{f})$ is given by

$$(1 - \lambda_2)^{**} = 1 - \lambda^{\circ}_{1}/\lambda \tag{14}$$

and
$$p_{int}^{**} = \frac{1 - \lambda^{\circ}}{\lambda_1(1 - \lambda_2)} p^{\circ}_{f} - \frac{c\,z\,N_1}{\lambda_1(1 - \lambda_2)L} - \frac{(1 - \lambda^{\circ})}{\lambda_1(1 - \lambda_2)} rK,$$

or
$$p_{int}^{**} = \frac{1 - \lambda^{\circ}}{\lambda_1(1 - \lambda_2)}(p^{\circ}_{f} - rK) - \frac{c\,Z}{\lambda_1(1 - \lambda_2)L} \tag{15}$$

where $Z = z \cdot N_1$ and equals total customer capacity provided by stage 1 firms in the case of partial vertical integration.

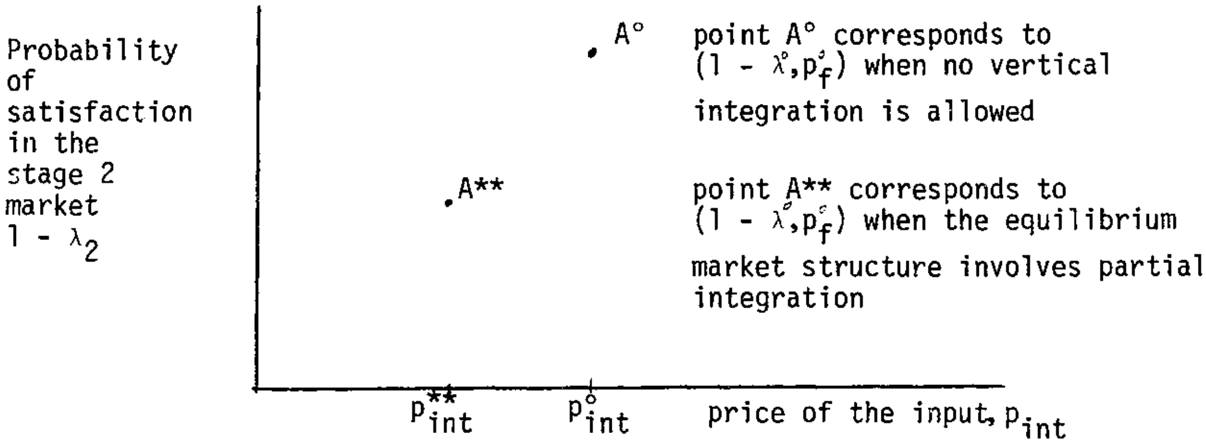
From Lemma 2, we know that wherever A** lies, the slope of the relevant derived isoutility curve at A** is less than that at A°. From (12) it is clear that point A** lies below point A° (see Figure 5-9). This last result makes intuitive sense since it says that to satisfy $1 - \lambda^{\circ}$ of the customers, you can satisfy less than $1 - \lambda^{\circ}$ in the stage 2 market provided you satisfy some customers in the stage 1 market by using those stage 1 firms' supplies of the input.

It is impossible to tell from (15) whether A** lies to the right or left of A°. It will depend on the relative magnitude of the terms in (15).

[1] The functional relation between $1 - \lambda_1$ and z was derived in Appendix A.

The diagram below depicts the situation where A** lies to the left of A°.

Figure 5-9 Derived Isoutility Curves and Vertical Integration



Suppose now that point A° is the market equilibrium point for the case of no vertical integration, so that all points with a superscript "o" now represent equilibrium quantities. Let us see if, in this case, we can determine if point A° lies to the left or right of point A**, which corresponds to the same $(1 - \lambda^\circ,\ p_f^\circ)$ as point A°. Rewrite (15) as

$$p_{int}^{**} = \frac{(1 - \lambda^\circ)}{\lambda_1(1 - \lambda_2)}\, p_{int}^\circ - \frac{c\ Z}{\lambda_1(1 - \lambda_2)L}, \tag{16}$$

where $\lambda^\circ = \lambda_1 \cdot \lambda_2$. We wish to determine if p_{int}^{**} is greater or less than p_{int}°. Since point A° represents the market equilibrium inthe case of no vertical integration, we know from the zero profit condition for stage 2 firms that

$$(1 - \lambda^\circ)\ p_{int}^\circ\ L = N_2 \cdot s_2^\circ \cdot c = S_2^\circ \cdot c,$$

where S_2° equals the total customer capacity provided by all stage 2 firms in the equilibrium when no vertical integration is allowed. Substituting this zero profit expression into (16), we obtain

$$p_{int}^{**} = p_{int}^{\circ} \left[(1 - \lambda^{\circ}) - \frac{(1 - \lambda^{\circ})\, Z}{S_2^{\circ}}\right] \frac{1}{\lambda_1 (1 - \lambda_2)},$$

or
$$p_{int}^{**} = p_{int}^{\circ} \frac{1 - \lambda^{\circ}}{\lambda_1 - \lambda^{\circ}} \cdot \left[1 - \frac{Z}{S_2^{\circ}}\right],$$

or
$$p_{int}^{**} = p_{int}^{\circ} R(Z),$$

where
$$R(Z) = \frac{1 - \lambda^{\circ}}{\lambda_1 - \lambda^{\circ}} \left[1 - \frac{Z}{S_2^{\circ}}\right].$$ [1]

To see if point A** lies to the right or left of A° we need to determine if R(Z) is greater or less than 1. Notice that $R(Z) < 1$ when $\frac{1 - \lambda^{\circ}}{\lambda_1 - \lambda^{\circ}} \left[1 - \frac{Z}{S_2^{\circ}}\right] < 1$. If, in the equilibrium market structure, the total customer capacity Z that stage 1 firms provide for themselves exceeds the total customer capacity, S_2°, that occurs when vertical integration is not allowed, then R(Z) always is below 1, and therefore point A** lies to the left of point A° in the previous figure.[2] For R(Z) to be less than 1 when $Z < S_2^{\circ}$, we require

$$\left[\frac{1 - \lambda^{\circ}}{\lambda_1 - \lambda^{\circ}} - 1\right] S_2^{\circ} < \frac{1 - \lambda^{\circ}}{\lambda_1 - \lambda^{\circ}} Z$$

or
$$(1 - \lambda_1)\, S_2^{\circ} < (1 - \lambda^{\circ})\, Z$$

or
$$\frac{(1 - \lambda_1)}{Z} < \frac{(1 - \lambda^{\circ})}{S_2^{\circ}} \tag{17}$$

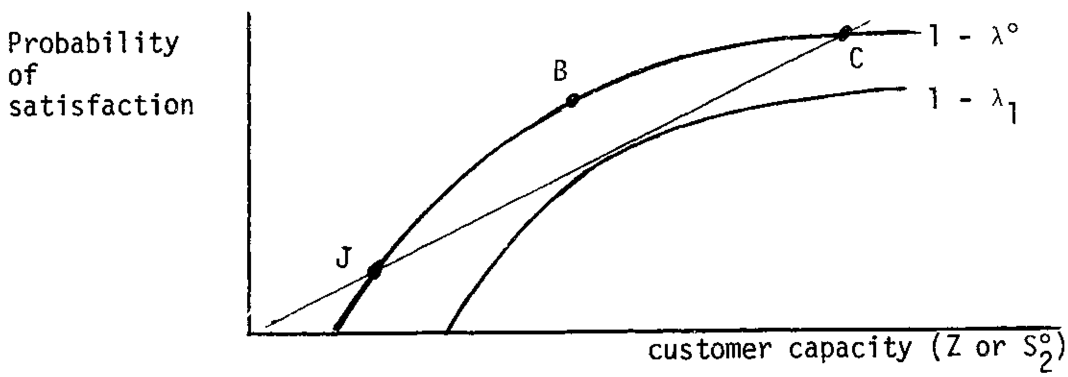
---

[1] Recall that $1 - \lambda^{\circ}(s_2, N_2)$ and $1 - \lambda_1 = 1 - \lambda_1(z, N_1)$, where the functional form for $1 - \lambda$ was derived in Appendix A.

[2] In fact, for this case, point A* lies in the second quadrant in Figure 5-9.

The term on each side of (17) is the probability of satisfaction per unit of customer capacity. The left hand term refers to the case involving partial vertical integration, while the term on the right refers to the case involving no vertical integration. Since stage 1 firms are less efficient absorbers of risk than stage 2 firms, we know that the function $1 - \lambda_1$ lies entirely below the $1 - \lambda^{\circ}$ curve as depicted below.[1] The terms on each side of (17) are the slopes of the rays from the origin to points on each curve.

Figure 5-10 Probability of Satisfaction



Whether inequality (17) will hold will depend on the equilibrium point in the case of no integration, $1 - \lambda^{\circ}$ and S_2°, and on the market equilibrium and Z associated with the equilibrium market structure involving partial integration. If we measure the degree of vertical integration that occurs in the equilibrium market structure involving partial vertical integration by the ratio of Z to S_2° (i.e. the ratio of the total customer capacity provided by stage 1 firms themselves in the case of partial integration to the total customer capacity provided by stage 2 firms in the case of no vertical integration), then from Figure 5-10, it is evident that

[1]This fact, and the fact that the curves are concave follows from Appendix A.

when the degree of vertical integration is either large (1 or above) or small (near 0), the inequality will hold.

From Figure 5-10, it is easy to see that if the market equilibrium associated with point A° lies near the unitary elastic point of the $1 - \lambda^\circ$ curve (i.e. anywhere along the arc JBC in Figure 5-10), then the inequality (17) will hold. More importantly, the inequality becomes more and more likely to hold as the $1 - \lambda_1$ curve falls further below the $1 - \lambda^\circ$ curve. In other words, the greater the difference in the ability of the stage 1 and stage 2 firms to absorb risk (i.e. the greater the difference between N_1 and N_2) the more likely is the inequality to hold.

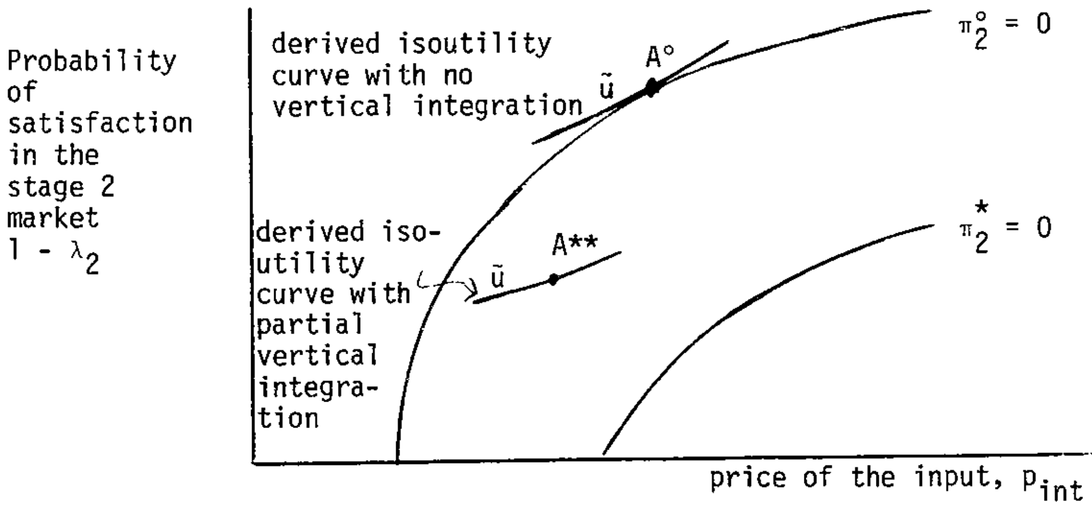
The case of most interest to a policy maker is where N_1 is much greater than N_2, because in this case the stage 2 firms are much better absorbers of risk than the stage 1 firms, yet because of the private incentives for vertical integration, the stage 1 firms may produce some of their own input. The social costs of vertical integration grow more serious as the differential between the stage 1 and stage 2 firms to absorb risk increases. Since we expect (17) to hold in this case of most interest, we carry out the subsequent analysis under that assumption. For future reference, let us formalize this as:

<u>Assumption 4</u>: Inequality (17) holds, where S_2° is the total customer capacity provided by stage 2 firms in the equilibrium when no vertical integration is allowed, and Z is the total customer capacity provided by stage 1 firms for themselves in the equilibrium market structure involving partial vertical integration.

If Assumption 4 is true, then we know that point A** that corresponds to the same $(1 - \lambda^\circ, p_f^\circ)$ as point A°, lies below A° and to the left of A°.

Also, from Lemma 2, the slope of the relevant isoutility curve is lower at A** than at A°. These results are depicted in the figure below.

Figure 5-11 Derived Isoutility Curves



In the diagram $\pi_2^\circ = 0$ is the zero profit curve of the stage 2 firms for the case of no vertical integration. It is drawn on the assumption that there are L consumers and N_2 firms. In the case involving partial vertical integration, equilibrium is determined by the tangency between the derived isoutility curves and the $\pi_2^* = 0$ curve. As discussed in Section 5.4, the $\pi_2^* = 0$ curve is drawn on the assumption that there are N_2 firms and L^e customers, where L^e equals the expected number of customers that stage 2 firms will see. Since in the case of partial vertical integration, some consumers are satisfied by stage 1 firms, L^e is less than L. Therefore, from the results of Chapter 2 the $\pi_2^* = 0$ curve lies below the $\pi_2^\circ = 0$ curve.

The final assumption we impose is on the shape of the derived isoutility curves. Recall from Chapter 2 that it is plausible to expect that the isoutility curves will be concave in $(1 - \lambda, p_f)$ space over the regions of interest. Moreover, from Assumption 1, the slope of these isoutility curves falls for any fixed p_f as $1 - \lambda$ falls. Since market equilibrium with partial vertical integration is determined in the stage 2 market using derived isoutility curves, it is necessary to make the stronger assumption:

Assumption 5: The derived isoutility curves are concave in $(1 - \lambda_2, p_{int})$ space over the relevant range and Assumption 1 applies to these derived isoutility curves over the relevant range.

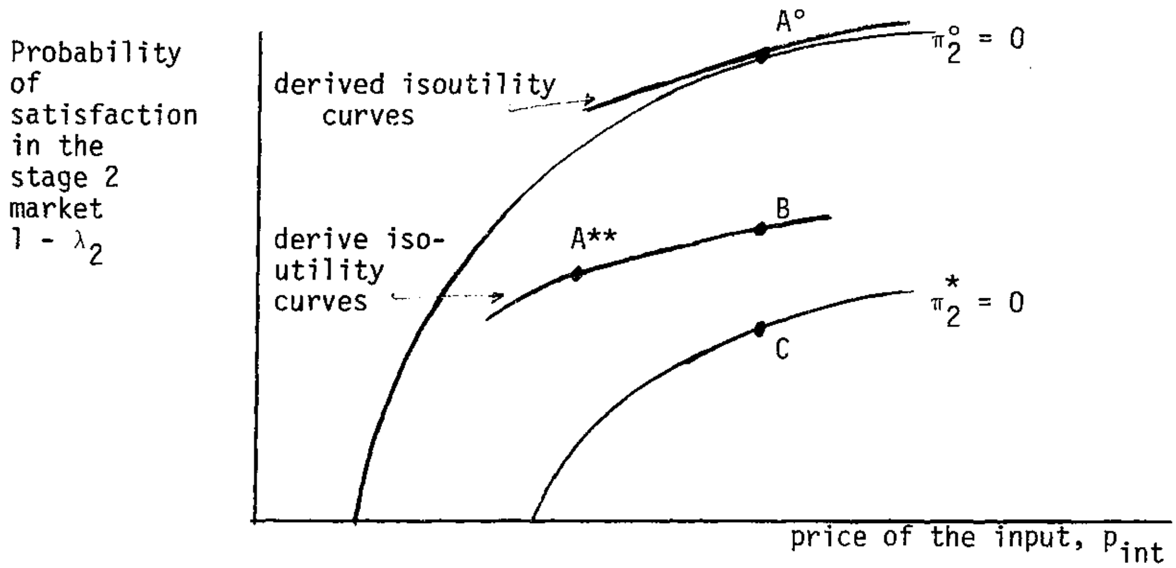
We are assuming that the isoutility curve that defines equilibrium in the case of no vertical integration is sufficiently concave beyond point A° (see previous diagram) so that the corresponding isoutility curve in the case of partial vertical integration is concave beyond point A**.

We are now ready to compare the equilibrium quantities that result for the equilibrium market structure involving partial vertical integration to the equilibrium quantities that would result if no vertical integration were allowed.

Theorem 3: Under Assumptions 1, 2, 4, 5, the equilibrium price in the factor market is higher in the market structure involving partial vertical integration than in the one involving no vertical integration.

Proof: Equilibrium for the market structure involving no vertical integration is depicted below at point A°, the tangency between the $\pi_2^\circ(L,N_2) = 0$ curve (i.e. the zero profit curve for stage 2 firms) and the relevant derived isoutility curve. For the case involving partial vertical integration, equilibrium is determined as the tangency between the relevant derived isoutility curve and the $\pi_2^*(L^e,N_2) = 0$ curve, where L^e is less than L.

Figure 5-12 Equilibrium With and Without Partial Vertical Integration



Consider how the derived isoutility curves shift in the case of vertical integration. From Assumption 4, the point A**, that corresponds to the same $(1 - \lambda^\circ, p_f^\circ)$ as point A°, lies to the southwest of point A°. Letting $m_u(K)$ stand for the slope of the relevant isoutility curve at point K, we have that from Lemma 2

$$m_u(A^\circ) > m_u(A^{**}).$$

Consider the derived isoutility curve on which point A** lies. From Assumption 5, the slope along this curve declines as p_{int} increases. Hence,

$$m_u(A^{**}) > m_u(B),$$

where point B lies directly below point A°. (See Figure 5-12.)

Consider the slope of the derived isoutility curve at point C (see Figure 5-12) which lies directly below point B and lies on the $\pi_2^* = 0$ curve. From Assumptions 1 and 5, we have that

$$m_u(B) > m_u(C).$$

Now, consider the slope along the $\pi_2^\circ = 0$ curve at point A°. From the definition of equilibrium, we have

$$m_u(A^\circ) = m_{\pi_2^\circ = 0}(A^\circ),$$

where $m_{\pi_2^\circ = 0}(A^\circ)$ is the slope of the zero profit curve, $\pi_2^\circ(L,N_2) = 0$, at point A°. The $\pi_2^\circ = 0$ curve is drawn on the premise that there are L consumers, while, because of the vertical integration, the $\pi_2^* = 0$ curve is drawn on the premise that there are less than L consumers in the stage 2 market. It follows from the theorems in Appendix 2 that

$$m_{\pi_2^* = 0}(C) > m_{\pi_2^\circ = 0}(A^\circ).$$

Combining all these inequalities, we obtain that

$$m_{\pi_2^* = 0}(C) > m_u(C),$$

or equivalently that at point C the slope of the isoutility curve is flatter than that of the zero profit curve. From this last inequality and Assumption 2, it follows that p_{int} is higher in the market equilibrium involving partial vertical integration than in the market equilibrium that would result if vertical integration were not allowed. Q.E.D.

Theorem 4 proved that the price of the factor, p_{int}^*, in the market equilibrium with partial vertical integration, exceeds the price, p_{int}°, in the market equilibrium with no vertical integration. Can we say how the price p_f of the final good compares between the two situations?

Using (12), we find that in the case of no vertical integration,

$$p_f^\circ = p_{int}^\circ + rK. \tag{18}$$

Using (15), we find that in the case of partial vertical integration,

$$p_f^* = \frac{\lambda_1(1 - \lambda_2)}{(1 - \lambda)} p_{int}^* + \frac{c\ Z}{(1 - \lambda)L} + rK.$$

In the equilibrium involving partial vertical integration stage 2 firms earn zero profits or

$$(1 - \lambda_2)\lambda_1 L\ p_{int}^* - c \cdot S = 0,$$

where S = total customer capacity provided by stage 2 firms in the partially integrated equilibrium. Substituting this zero profit condition into the previous equation, we obtain

$$p_f^* = \frac{\lambda_1(1 - \lambda_2)}{(1 - \lambda)} [1 + \frac{Z}{S}]\ p_{int}^* + rK,$$

or $$p_f^* = b\ p_{int}^* + rK, \qquad (19)$$

where $$b = \frac{\lambda_1(1 - \lambda_2)}{(1 - \lambda)} (1 + \frac{Z}{S}).$$ [1]

Comparing (18) and (19), we see that p_f^* will exceed p_f° if $b \cdot p_{int}^*$ exceeds p_{int}°. From Theorem 3, we know that $p_{int}^* > p_{int}^\circ$. However, there appears to be no compelling reason why b should exceed 1. Therefore it is impossible using (19) to determine whether p_f^* exceeds p_f° without making further assumptions. It seems possible, though I suspect unlikely, that with partial integration, the price of the stage 2 input could rise, but

[1] The reader should be careful not to confuse the variables in this expression for b with those in (17). In the expression for b, all variables refer to their values in the market equilibrium involving partial vertical integration. In (17), the variables $1 - \lambda^\circ$, and S_2° refer to values in the market equilibrium that results when no vertical integration is allowed.

the price of the final good could fall.[1] In this case, we know from Theorem 1 that the probability of satisfaction would have to fall sufficiently so that consumers are worse off in the case of partial vertical integration than in the case of no vertical integration.

5.9 Market Structure and The Choice of the Output Technology

In this section we examine how the transmission of uncertainty between firms can influence the choice of technology. So far, we have assumed that there is only one technology to produce the output, namely a Leontief technology which uses K units of capital and one unit of the input subject to shortages to produce one unit of output. Now we will assume that there suddenly becomes available a new Leontief technology with input requirements (K_1,ℓ). We examine the incentives for introduction of the new technology in a nonintegrated and integrated market setting. The main conclusion of this section is that introduction of a new technology that would benefit society is more likely to occur in a market with vertical integration than in one without vertical integration.

In the discussion of market clearing, it was seen that the input-output coefficients (K,1) influence the shape of the derived isoutility curve through (12) or (15). Recall from Secton 5.4 that the derived isoutility curves reflect the tradeoffs between the probability of obtaining the input, and the price of the input (i.e. the tradeoffs in $(1 - \lambda_2, p_{int})$ space)

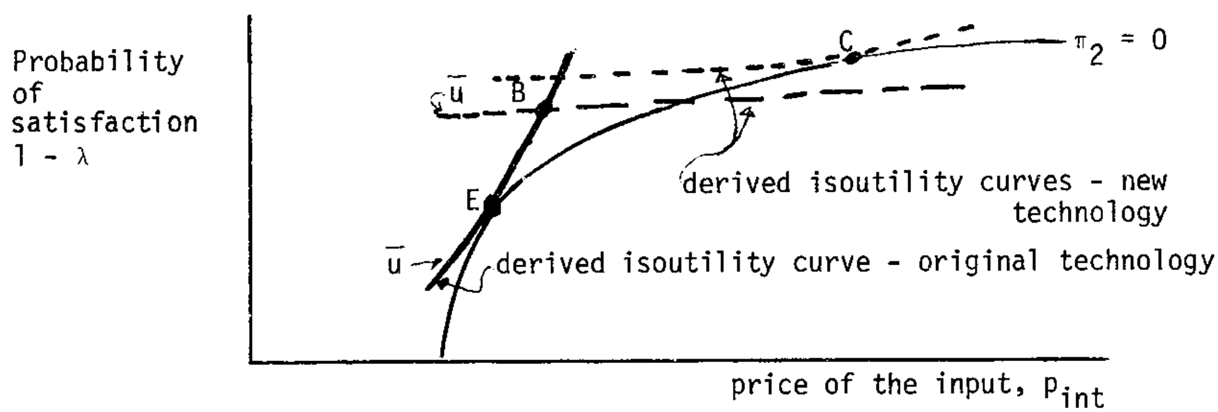
---

[1]This statement should not be construed as implying that I have found an example illustrating this possibility. As mentioned earlier, dealing with equilibrium market structures is analytically very complicated. However, the proofs of the previous theorems do not suggest any reasons why the possibility of $p_f^* < p_f^o$ cannot occur. It is of interest to note that when $p_f^* < p_f^o$ and hence $b < 1$, a stage 1 firm earns a negative per unit profit when it sells a final good made with an input purchased in the stage 2 market.

that translate into tradeoffs between the probability of obtaining the final good and the price of the final good (i.e. the tradeoffs in $(1 - \lambda, p_{int})$ space) that consumers are willing to make. So, for example, if we had a new technology (K_1, ℓ) where $\ell < 1$, any increase in p_{int} would translate through either (12) or (15) into a smaller price increase in p_f than it would if $\ell = 1$. In this case, the derived isoutility curves would become flatter than they are in the case of $\ell = 1$. We will refer to the original (K,1) technology as the "original" one, and the (K_1, ℓ) technology as the "new" one.

First, consider the market equilibrium that would occur if vertical integration is not allowed. Let the equilibrium factor price be p°_{int}. The stage 1 firms will adopt the new technology <u>only if</u> it is more efficient when the price of capital is r and the price of the input is p°_{int}. But this marginal calculation is not sufficient to guarantee that consumers would not be better off under the new technology. The diagram below illustrates this point.

<u>Figure 5-13 Choice of Technology</u>



Point E is the original market equilibrium. The derived isoutility curve through point E is drawn using the input-output coefficients of the old technology. The level of utility achieved by consumer along this curve is $\bar{u}$. The derived isoutility curve, corresponding to the same $\bar{u}$ when the input output coefficients of the new technology are used, is drawn as a dotted line. The fact that the dotted curve passes above point E is equivalent to the statement that the new technology is less efficient than the old technology at the factor price associated with point E. The two derived isoutility curves cross at point B, where each technology is equally efficient (i.e. $p^B_{int} + rK = p^B_{int} + rK_1$). Beyond point B, the new technology is more efficient. Notice that the dotted isoutility curve crosses the zero profit ($\pi_2 = 0$) curve. Therefore, there exists some point C of tangency between the derived isoutility curve with the new technology and the $\pi = 0$ curve that represents a level of utility above $\bar{u}$.[1]

Consumers would be better off if all stage 1 firms adopted the new technology so that the market equilibrium would move to point C. Yet, because stage 1 firms have no control over the input market, they will not have any incentive to adopt the new technology, since it is inefficient at the initial market equilibrium E. The existing prices do not provide incentives for stage 1 firms to change technologies, nor for stage 2 firms to alter their behavior.

It is easy to see how vertical integration could remedy this situation. To make the point, it suffices to consider the case of complete vertical integration. Since each stage 1 firm totally controls its production of the input, it can coordinate its $(p_{int}, 1 - \lambda)$ mix[2] to its own specifications.

---

[1]The new technology does not alter the zero profit (i.e. $\pi = 0$) curve for stage 2 firms. It only affects the per customer demand for the input. Recall from Appendix A that the $\pi = 0$ curve is independent of the per capita demand.

[2]Recall that in the case of complete vertical integration, the stage 1 firms effectively act as stage 2 firms since they produce their own input.

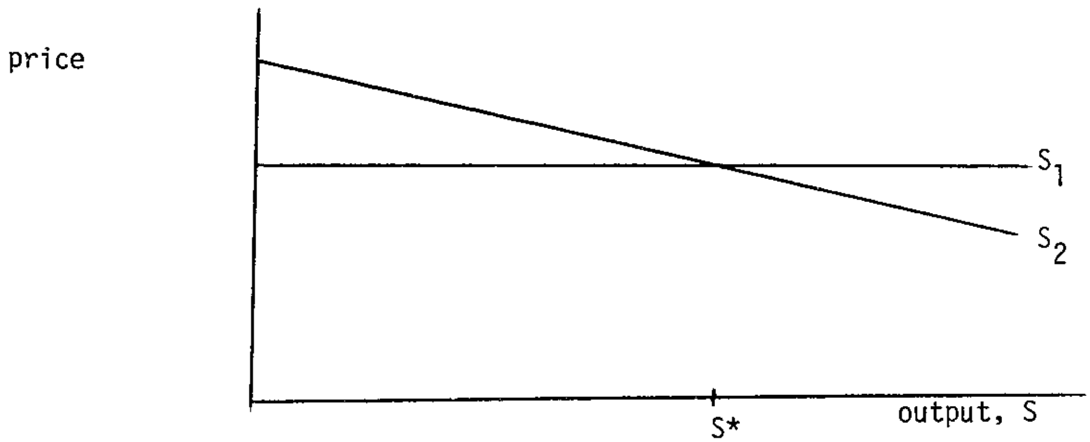
Because of this possibility of coordination, stage 1 firms will be able (and through competition will be forced) to move immediately to any achievable point that justifies the use of the new technology and makes consumers better off.

It is precisely because vertically integrated firms can exercise control over the characteristics ($p_{int}, 1 - \lambda$) of the input, that they are able to introduce the new and more desirable technology. With no vertical integration, price signals are not sufficient to convey the benefits of switching to a new technology. The basic reason is that whether consumers would be better off if all stage 1 firms adopted a new technology is a nonmarginal change. In a competitive market, the decision of what technology an individual should adopt is a marginal change. In this case, marginal incentives at the firm level do not give the correct signals as to whether the nonmarginal change is desired.

A similar problem about choice of technology arises in relation to Marshall's concept of a competitive long run supply curve that slopes downward because of pecuniary economies. Because of a lowering in input prices that accompanies output expansion, Marshall argued that long run supply curves could slope downwards. Behind every supply curve lies a production technology, yet it is not made very clear how the production technology changes along this downward sloping supply curve. If there is only one technology, then no problem arises. But suppose that there are two technologies; one which leads to a horizontal long run market supply curve S_1 and one which leads to a downward sloping one, S_2. (See figure below).

Supply curve S_2 is drawn on the assumption that all firms use technology 2 so that economies of scale in the production of some factor can occur as industry output expands. For $S < S^*$, it is better for all stage 1 firms to

Figure 5-14 Choice of Technology with Downward Sloping Supply Curves



adopt technology 1, while for S > S*, it is better for them all to adopt technology 2. Yet, if the firms were originally using technology 1, there may well never be an incentive for an individual firm to switch to technology 2. The economies of scale in production of the input (which causes the S_2 curve to slope downward) will not be fully realized if only one firm adopts the new technology. In this case, the marginal calculation of an individual firm to adopt the new technology will not coincide with the nonmarginal calculation of whether all firms should adopt the new technology.

The point of the above discussion is that when the choice of technology affects the equilibrium characteristics (e.g. price, probability of satisfaction) of the input which, in turn, influence the choice of technology, then it is not necessarily true that the individual decisions by firms will lead to the correct technology being adopted. In such cases, existing prices need not provide the correct signals for choice of a new technology. In the model of this chapter, vertical integration is a mechanism by which individual firms can gain control over the two characteristics (i.e. price and probability of satisfaction) of the input that influences its choice of output technology. The final product firm that is vertically integrated is thereby better able to coordinate its choice of output technology with the characteristics of the input. With the flexibility of tailoring the input characteristics to its needs, the vertically integrated firm may introduce the new more desirable technology, while the nonintegrated firms who must take the input characteristics as given in the marketplace may get locked into the old technology and have no incentive to change production technologies.

In markets characterized by uncertainty, Theorem 1 proved that vertical integration can be socially undesirable. A more efficient market structure will usually involve no vertical integration. However, new and socially desirable technologies are more likely to be developed and introduced in a market structure involving vertical integration in which individual firms can coordinate input characteristics with their choice of technology, than in a market structure involving no vertical integration in which such individual firm coordination is impossible. We are led to the Schumpeterian view of the world that it may be necessary to tolerate some static inefficiency in market structure in order to create an environment in which new and socially desirable technologies can be developed and introduced.

5.10 Policies Toward Potential Vertical Integration In A Vertically Integrated Market

The model has an interesting implication as regards policies toward vertical integration through internal growth by new entrants or by non-integrated firms in a market structure where vertical integration is already present. The current government policy would seem to be that initially vertical integration is allowed; however, as the number of vertically integrated firms increases the government becomes increasingly reluctant to allow any additional vertical integration to occur.

So, imagine a market structure that already has some vertical integration. The vertical integration could arise as described in Section 5.7, or else the industry could be very young and be characterized by firms "born integrated" because of the nonexistence of a market for the input supply.[1] For such cases, it is not true that preventing an additional firm from vertically integrating through internal growth is the socially desirable course of action. In this third best[2] world involving a competitive market structure and vertical integration, it could well be more efficient to allow the additional firms to vertically integrate through internal growth. To see this point simply, if a number of stage 1 firms produce a significant amount of their own input, then the effective customer per store ratio that stage 2 firms see could be very small in comparison to the customer per firm ratio that nonintegrated stage 1 firms see, even though the number of stage 2 firms is less than the number of stage 1 firms. In this case, the nonintegrated stage 1 firms may initially be better risk absorbers than the stage 2 firms. Hence, some vertical integration for a nonintegrated stage 1 firm would be both privately and socially desirable. In general, the greater the amount of vertical integration initially in the market, the larger is the social loss that would occur from prohibiting a nonintegrated stage 1 firm from vertically integrating through internal growth.

[1]G. Stigler, "The Division of Labor is Limited by the Size of the Market," Journal of Political Economy, 1951, p. 188, and J. Green, p.21, op.cit.

[2]From Theorem 1 it follows that it would be better to force all firms to become nonintegrated. Even better would be to provide lump sum transfers to stage 2 firms in this nonintegrated market structure. This last statement follows immediately from the results of the previous chapter.

5.11 Growth vs. Merger

A recurring issue in the literature on vertical integration has been whether internal growth or mergers with suppliers is the socially more desirable method of vertical integration. "There is a different public policy standard for vertical integration achieved through merger rather than by internal growth."[1] Antitrust policy has reflected the view that internal growth is not harmful, while mergers are. Proponents of this position rely on the "foreclosure" theory which argues that mergers limit the ability of nonintegrated firms to purchase supplies while internal growth does not.[2] Opponents of the foreclosure theory argue that it is not integration that is at fault, but monopoly power.[3]

So far, the model has investigated the effects of vertical integration through internal growth. In this section, the effects of vertical integration through merger are considered. We examine the internal growth vs. merger issue in the context of the model presented earlier. The issues of growth and merger are examined from the viewpoint of their different effects on the transmission of uncertainty between firms. We will not investigate questions associated with increasing monopoly power, and instead will continue to assume that the firms compete with each other in the manner described in Chapter 2.

As before, we assume that the number of stage 2 firms is less than the number of stage 1 final product firms. Since each stage 2 firm has constant returns to scale, we must make some assumption about fixed capacity in

---

[1] W. Mueller, "Public Policies Toward Vertical Mergers," in J. Weston and S. Peltzman, op.cit.

[2] P. Areeda, "Structure Performance Assumptions in Recent Merger Cases," in ibid.

[3] R. Bork, "Vertical Integration and Competitive Processes," in ibid.

order to make sense of a foreclosure theory. So suppose that with no vertical integration, the market has reached equilibrium, and that the capacity of the stage 2 firms is fixed at the production level associated with this equilibrium. Each stage 2 firm supplies input for several stage 1 firms, so that if some stage 1 firm got control of a stage 2 firm, some of the supply source could be foreclosed to other stage 1 firms.

By merger, we mean that one stage 1 firm gains control of one stage 2 firm which then produces exclusively for the stage 1 firm.[1] If one stage 1 firm merges, then it will have a competitive advantage over the other stage 1 firms, and will force these other stage 1 firms to try to merge. Since there are N_1 stage 1 firms, and only N_2 stage 2 firms, we know that only N_2 of the stage 1 firms will be successful in a merger attempt. So, if mergers occur, $N_1 - N_2$ of the stage 1 firms will be forced out of business, because they will be unable to obtain supplies of their input. When this happens, the resulting market equilibrium with the N_2 merged stage 1-stage 2 firms that remain is identical to the market equilibrium in the case of no vertical integration.[2]

Internal growth means, as before, that stage 1 firms produce their own input and possibly use the stage 2 markets as (insurance-like) markets for supplying the input. Internal growth, then, expands the production capacity of the input. With internal growth, each of the N_1 stage 1 firms will produce some of the input for itself. From Theorem 1, we know that, because of the transmission of uncertainty between the stage 1 and stage 2 markets,

---

[1]This situation can equivalently be regarded as forward integration with each stage 2 firm obtaining control of one stage 1 firm.

[2]This assumes that the utility of the consumer depends only on the price and probability of satisfaction and not on the number of firms.

internal growth leaves consumers worse off than they were in the case of no vertical integration.

It then follows that if vertical integration is to occur, it is preferable to have it occur through merger and not internal growth. For the model under study, it is more desirable to get rid of some of the stage 1 firms than to allow stage 1 firms to expand internally, and thereby impose costs on society by virtue of their less efficient risk absorbing capacities. We see then that when we focus on the effects caused by the transmission of uncertainty between consumers, final product, and input product firms, we reach a conclusion about growth vs. merger that is diametrically opposite that held by the proponents of the foreclosure theory.

5.12 Growth vs. Pre-purchased Contracts

In the previous section, we assumed that vertical integration through merger meant that some stage 1 firms would be forced to go out of business for lack of a source of supply. In this section, we consider another type of merger, one through pre-purchased contracts. In merger by pre-purchased contract,stage 2 firms sign contracts with stage 1 firms to deliver at constant cost c some amount of input which the stage 1 firms must receive at the beginning of each market period. As in the previous section, we assume that the capacity of the stage 2 firms is fixed so as to give the foreclosure theory meaning in the model.

In merger by contract, each of the N_1 stage 1 firms is able to tie up some fraction of stage 2 capacity. In contrast to merger by acquisition, all N_1 stage 1 firms are able to remain in the market. Internal growth means that stage 1 firms create additional capacity to produce the input for themselves.

The equilibrium market structure with internal growth is defined as before. With merger by pre-purchased contract, we see that provided the fixed capacity of the N_2 stage 2 firms is not reached, the market structure will be identical to that achieved with internal growth. Let us use the total customer capacity provided by stage 1 firms for themselves (either through internal growth or pre-purchased contracts) as a measure of the amount of vertical integration that occurs. If the capacity constraint of stage 2 firms is reached and if the degree of vertical integration is unchanged from the internal growth equilibrium market structure, then consumers will be better off with vertical integration occurring through internal growth. Heuristically, this occurs because when the fixed capacity becomes a constraint, consumers desire more output which can only be produced if capacity is expanded through internal growth. More technically, when the fixed capacity becomes a constraint, the tangency between the derived isoutility curves and the zero profit curve occurs at an unattainable point on the zero profit curve. With capcity expansion through internal growth, this previously unattainable point can be reached (or at least approached).

On the other hand, in the integration by pre-purchased case, the fixity of stage 2 capacity might cause the degree of vertical integration that occurs in the equilibrium market structure to differ from that in the integration by internal growth case. If this is the case, then it is not possible to say whether consumers are worse off when vertical integration occurs through internal growth or when it occurs through pre-purchased contracts. All we can say is that consumers will tend to be better off in the integration through contract case when the amount of vertical integration is lower than it is in the case of integration through internal growth. In any case, it follows from the previous section, that vertical integration through merger is superior to vertical integration through either internal growth or pre-purchased contracts for the situations under study.

5.13 Forward Integration

The model of this chapter was designed to focus on the incentives, caused by the transmission of uncertainty, for final product firms to integrate backward by producing their own input. It is also possible within this model to give an interpretation to forward integration.

We regard forward integration as occuring when a stage 2 input firm gains control of one or more stage 1 final product firms, which then only buy inputs from the stage 2 firm. As before we assume that (for cost reasons) stage 1 firms cannot ship the input amongst themselves. If each stage 2 firm owns N_1/N_2 (assume that N_1 is divisible by N_2) stage 1 firms, then the market equilibrium is identical to that for the case of no integration. Each stage 2 firm will realize that it is unprofitable for it to ever allow any of its stage 1 firms to produce any input for itself, and so all the inputs produced will be held at the stage 2 level until a stage 1 firm announces an input demand. Forward integration, just like backward integration through merger, leads to the market equilibrium that is identical to that of the socially preferred market structure of no vertical integration, by creating incentives so that only the N_2 stage 2 firms produce the input and bear the risk of using it.

5.14 Horizontal Merger and Social Welfare

Since the number of stage 1 and stage 2 firms is exogenous to the model, one has to be very cautious about welfare interpretations when the number of firms change. One important caveat associated with the model that has been presented deals with the implications of horizontal merger. In the model, it appears that the utility of consumers increases as the number of stage 2 firms declines to one. One has to be careful to avoid the inference that for the markets under study total horizontal integration of stage 2 firms is always desirable. The model is designed to study the transmission of uncertainty between firms in a competitive environment. However, large horizontal mergers could create monopoly power which by itself entails social costs. Horizontal integration may be desirable from the point of view of risk sharing (which is what the model is designed to focus on), but undesirable from the point of view of creating monopoly power. (Whether there are any private incentives for horizontal merger to occur is an altogether different question which was already addressed in Section 2.10 of Chapter 2.)

The results of this chapter indicate that, when the fixed number of stage 1 firms exceeds that of stage 2 firms, a nonintegrated competitive market structure is socially preferred to an integrated one. However, it is not true, for a competitive market with the given number of stage 1 and stage 2 firms, that the nonintegrated competitive market structure is socially optimal. It follows from the previous chapter that the socially optimal solution in this case is to have a nonintegrated market structure and usually to pay lump sum subsidies to the stage 2 firms. These subsidies would usually be used to encourage the stage 2 firms to expand production of the input that is subject to shortages.

5.15 Demanders of Stage 2 Inputs

So far, the model has assumed that the only demanders of the stage 2 input are the stage 1 firms. How is the previous analysis affected if there are other segments of the economy that demand the stage 2 input?

Suppose first that the per capita input demand is the same for all customers who demand the input. The effect of having additional stage 2 customers is to increase the risk absorbing capability of the stage 2 firms over that of the stage 1 firms. Let L and L* stand respectively for the number of stage 1 customers and the number of stage 2 customers other than those traceable to stage 1. As before, let N_1 and N_2 stand respectively for the number of stage 1 and stage 2 firms. Then, the stage 2 firms will be the more efficient risk absorbers if $\frac{L + L^*}{N_2} < L/N_1$. When there are stage 2 customers who do not come from stage 1, we see that it is not necessary for N_2 to be less than N_1 for the stage 2 firms to be better absorbers of risk than the stage 1 firms.

In terms of the model, as L* increases, the price-probability of satisfaction combinations that stage 2 firms can achieve improves (i.e. the $\pi_2 = 0$ curve shifts up), and the incentives for vertical integration by stage 1 firms is reduced. Indeed, for sufficiently large L*, we know from the theorems of Chapter 2 that the price of the stage 2 input, p_{int}, will approach its cost of production, c, and the probability of satisfaction will approach 1, so that the incentive for vertical integration, expressed in (4), will not occur for any given number, N_1, of stage 1 firms.

On the other hand, sharing the stage 2 market with other customers can present a problem to the stage 1 firms. As seen in Chapter 2, if segmented markets do not develop then the tastes of the majority between price and probability of satisfaction will influence equilibrium. If the stage 1 customers have drastically different tastes from the majority, then incentives may be created for the stage 1 firms to produce their own input so as to achieve a price-probability of satisfaction combination that they would find preferable.

For the markets under study, price compensates firms for the average risk in the market. If there are different classes of customers all participating in same market and if one class of customers has a very uncertain per capita demand each period, the price of the good is driven up and is paid for by the other "less risky" customers.[1] In such cases, the stage 1 customers will have an incentive to produce the input for themselves and thereby avoid having to pay for the costs that someone else's "riskiness" of demand imposes.

We see then that when there are customers other than those traceable to stage 1 firms demanding the input, that the additional incentives for stage 1 firms to produce the input for themselves are offsetting. Different tastes for price and probability of shortages, and different "riskiness" of demands among the customers can create incentives for stage 1 to vertically integrate. The increased number of stage 2 customers increases the differential in the risk absorbing ability between stage 2 and stage 1 firms, and creates disincentives for stage 1 firms to vertically integrate. When the number of "nonstage 1" customers is large, this latter affect is likely to predominate.[2]

[1] The classic example of such markets are those dealing with insurance.

[2] In fact, the strategy of Durant, when he was creating General Motors was to vertically integrate only in those factor markets where his demand accounted for a large part of the total demand. See Chandler, Chapter 3, op.cit.

5.16 Reinterpretation and Applications

Although the discussion of this chapter has dealt mainly with the incentives and consequences of firms to vertically integrate by producing some of their own inputs, many of the same arguments apply to the holding of different types of inventory among firms aligned in a stage of production sequence. To model that situation, we would want to allow for inventories to be held in either a final good or input good form, and for a lead time required for production at each stage of production. Additional complications could allow for varying delivery times between stages of production, and varying holding costs of inventory at the different stages of production. The question arises as to whether a free market which operates in the manner described in Chapter 2 will create incentives for firms to achieve a proper allocation of inventory holdings amongst themselves. Such questions which relate to the transmission of uncertainty between firms and to the proper allocation of risk amongst firms are similar to those discussed in this chapter.[1]

A straightforward interpretation of the model of this chapter is to regard the input subject to shortages as capital services. (The fact that in the model the input that was not subject to shortages was called "capital" is obviously irrelevant.) The model then provides a theory of how market structure and uncertainty influence the allocation of capital among firms, and how they affect the excess capacity and full employment of capital.

---

[1]There is however an important distinction between this inventory problem and the problem of vertical integration. In both problems, if a firm either purchases for inventory or produces for itself an extra unit of input, the firm raises the probability that it can satisfy its customers. However, in the vertical integration case, the firm also has the potential of saving $(p_{int} - c)$ by producing the input for itself at cost c rather than purchasing the input from someone else at price p_{int}.

Although the model always treated the input that is subject to shortages as producible, this is not a necessary feature of the model. The input could be regarded as land or labor, each of which is subject to shortages when bought on the free market. The model would then explain incentives of firms to "hoard" land or labor to insure themselves the ability to produce the output.[1] Just as in the case of vertical integration, there can be strong private incentives to "hoard", even though it is socially undesirable.

5.17 Summary

This chapter has presented a model of the transmission of uncertainty from the final product market to one of its factor markets. The demand uncertainty and price inflexibility discussed in Chapter 2 characterize market operation. For such markets, the assurance of adequate input supplies becomes a natural, and crucial, concern of firms. As seen in Chapter 2, the stochastic nature of demand affects the operating costs of firms. Since production of the input by stage 1 firms affects not only the mean level but also the entire stochastic structure of demand facing stage 2 firms, there is a type of externality in the model.

It was shown that free markets, if not interfered with, cannot be relied upon to achieve the optimal allocation of production of the input between the stage 1 final product firms and the stage 2 factor market firms. The free market equilibrium will not in general achieve the proper allocation of risk between stage 1 and stage 2 markets. Even when stage 1

---

[1] R. Hall, "An Aspect of the Economic Role of Unemployment," to appear in the Proceedings of the International Economics Association, April, 1975.

firms are less efficient absorbers of risk than stage 2 firms, strong private incentives to vertically integrate can exist. These incentives arise because a stage 1 firm can use its own input holdings to satisfy its high probability input demand and use the higher cost input from the stage 2 market to satisfy its low probability input demand.

Vertical integration through internal growth was shown to be socially undesirable, even though it can be privately desirable to stage 1 firms. The consequences of such vertical integration are a lower level of expected utility for consumers, and usually higher prices in both the final product and factor market. From the standpoint of static efficiency, the prohibition of vertical integration through internal growth is a desirable course of action for the circumstances studied in this chapter. On the other hand, from a more dynamic viewpoint, vertical integration may be more desirable. It was found that new and benefical technologies are more likely to be introduced in a market structure involving vertical integration than in one involving no vertical integration. In addition, it was shown that, in contrast to the thinking embodied in current antitrust policies, internal growth can be a much more harmful method of achieving vertical integration than a policy of mergers.

The results of this chapter emphasize the importance of distinguishing between market clearing under certainty and under uncertainty. An analyst using a deterministic approach to this problem would be led astray and would be unable to find any undesirable incentives or disincentives for vertical integration. It is only by explicitly analyzing the consequences of the uncertainty on market behavior that the undesirable incentives for and effects of vertical integration can be discovered.

APPENDIX C

Appendix to Chapter 5

Random Strategies

For the case of no and partial vertical integration, it was always assumed that if a stage 1 firm received a demand for its product, it would try to fill it by going into the factor market, if necessary. The stage 1 firm was never allowed the option of entering the stage 2 market only, say, k $(0 < k < 1)$ fraction of the time. For the case of no vertical integration, it is clear that any value other than $k = 1$ is nonoptimal, since for values of $k \neq 1$ the stage 1 firm is turning down profitable sales. In the case involving partial vertical integration, the argument is not so simple since, from the last footnote in Section 5.8, it appears possible that a stage 1 firm could actually lose money every time it sells a product using a stage 2 input. In this appendix, we show that in a market equilibrium involving no partial vertical integration, any randomized strategy is nonoptimal.[1]

Suppose that the market is in an equilibrium involving partial vertical integration as discussed in Section 5.4. Consider whether, without earning negative expected profits, a stage 1 firm could offer a higher level of utility to consumers if it entered the stage 2 market only k fraction of the time, when it ran out of its own inputs. If $1 - \lambda_2 =$ equilibrium probability of being satisfied in a stage 2 market,

[1]Even if this were not the case and randomized strategies were optimal in the model, there would be good reason to exclude such randomized strategies from the analysis. It is hard to imagine an actual firm following such a policy. More importantly, customers might find it undesirable to frequent a firm which they felt didn't try "very hard" to satisfy them (i.e. the value "k" of the randomized strategy could influence the utility level of the consumer).

then for this stage 1 firm, the effective probability of being satisfied if the firm has exhausted its own inputs is given by $k(1 - \lambda_2)$. Suppose that with this effective probability of satisfaction, $k(1 - \lambda_2)$ and the equilibrium price of the input, p_{int}, that the firm is able to earn non-negative expected profits and is able to offer a utility level greater than that offered in equilibrium. But from the shape of the zero profit ($\pi_2 = 0$) curve, we know that along the zero profit curve, the stage 2 firms could provide this same probability of satisfaction $k(1 - \lambda_2)$, but at a price p^*_{int} lower than p_{int}. If a stage 1 firm is able to provide a higher than equilibrium level of utility with the combination $((1 - \lambda_2)k, p_{int})$, it can surely do so with the combination $((1 - \lambda_2)k, p^*_{int})$ where $p^*_{int} < p_{int}$. Hence, if in equilibrium a randomized strategy exists that allows a stage 1 firm to offer a higher than equilibrium level of utility without earning negative expected profits, then along the $\pi_2 = 0$ curve, used to define the equilibrium, there exists a point other than the equilibrium point that also allows a higher than equilibrium level of utility to be achieved by a stage 1 firm that follows a nonrandomized ($k = 1$) strategy. But by definition of equilibrium, this is impossible and we obtain a contradiction.

Lemma 1: A stage 1 firm will continue to expand its customer capacity from z to z + 1, if

$$-c + \left[p_{int} + \lambda_2[p_f - \text{cost } 2]\right][1 - P(z)] > 0$$

where $P(z)$ = cumulative probability that fewer than z + 1 customers will visit a stage 1 firm, and all of the other previously defined variables are at their equilibrium values that result when each stage 1 firm produces sufficient input to satisfy z customers by itself.

Proof: The proof consists of straightforward algebra. The relation was defined previously in Chapter 5. Recall that cost 1 = rK + c, and cost 2 = rK + p_{int}.

Let $\pi(x/e)$ denote profits of a stage 1 firm with customer capacity x, when all values take on their equilibrium values e. From the definition of profits, it follows that

$$\pi(z/e) = \sum_{i=0}^{z} \Big[[p_f - \text{cost } 1]i - (z - i)c\Big]pr(i) \qquad \text{(C1)}$$

$$+ \sum_{i=z+1}^{\infty} \Big[[p_f - \text{cost } 1]z + [p_f - \text{cost } 2](i - z)(1 - \lambda_2)\Big] pr(i),$$

$$\pi(z/e) = A + B$$

where A and B refer to the two summation terms above. Because we are in equilibrium, we have that A + B = 0.

To obtain $\pi(z+1/e)$, substitute z + 1 for z in the above expression for profits (C1). After some manipulation, we obtain

$$\pi(z+1/e) = \sum_{i=0}^{z} \Big[[p_f - \text{cost } 1]i - ((z - i) + 1)c\Big]pr(i)$$

$$+ [p_f - \text{cost } 1](z + 1)pr(z + 1)$$

$$+ \sum_{i=z+2}^{L} \Big[[p_f - \text{cost } 1]z + [p_f - \text{cost } 1]$$

$$+ [p_f - \text{cost } 2][(i - z) - 1](1 - \lambda_2)\Big]pr(i),$$

or

$$= \sum_{0}^{z} \Big[[p_f - \text{cost } 1]i - (z - i)c\Big]pr(i) - \sum_{0}^{z} cpr(i)$$

$$+ [p_f - \text{cost } 1](z + 1)pr(z + 1) + \sum_{z+2}^{\infty} \Big[[p_f - \text{cost } 1]z$$

$$+ [p_f - \text{cost } 2](i - z)(1 - \lambda_2)\Big]pr(i) + \sum_{z+2}^{\infty} \Big[[p_f - \text{cost } 1]$$

$$- [p_f - \text{cost } 2](1 - \lambda_2)\Big] pr(i),$$

$$= A - \sum_{0}^{z} c\,pr(i) + [p_f - \text{cost } 1](z + 1)pr(z + 1) + B - \Big[[p_f - \text{cost } 1]z$$

$$+ [p_f - \text{cost } 2](1 - \lambda_2)\Big]pr(z + 1) + \sum_{z+2}^{\infty} \Big[[p_f - \text{cost } 1]$$

$$- [p_f - \text{cost } 2](1 - \lambda_2)\Big] pr(i),$$

$$= -c \sum_{i=0}^{z} pr(i) + [p_f - \text{cost } 1]pr(z + 1) + \sum_{z+2}^{\infty} \Big[[p_f - \text{cost } 1]$$

$$- [p_f - \text{cost } 2](1 - \lambda_2)\Big] pr(i) - [p_f - \text{cost } 2](1 - \lambda_2)pr(z + 1),$$

$$= -c \sum_{0}^{z} pr(i) + \sum_{z+1}^{L} \Big[[p_f - \text{cost } 1] - [p_f - \text{cost } 2](1 - \lambda_2)\Big]pr(i), \text{ or}$$

$$\pi(z+1/e) = -c + \sum_{z+1}^{L} \Big[p_{int} + \lambda_2[p_f - \text{cost } 2]\Big]pr(i) \qquad \text{Q.E.D.}$$

CHAPTER 6

Conclusion

This research has tried to develop insights into the behavior of markets characterized by demand uncertainty, price inflexibility, and a lead time required for production. In the investigation of vertical integration, supply uncertainty also became an essential feature of market operation. For the markets under study, a firm never knows how much of its product will be demanded during a market period. Prices respond to market forces but do not adjust at each instant of time to keep supply and demand in balance. In terms of the analysis, it does not matter whether there is a lead time in production, just as long as there is some prior committment to production that must be made before demand can be observed. It would appear that for many markets the description of the markets under study here is more applicable than is the traditional description of markets, where somehow price instantaneously adjusts to always keep supply and demand precisely in balance.

The essential feature of market operation deals with the risk caused by the demand (or supply) uncertainty. Demand uncertainty imposes costs on a firm. The entire stochastic structure of demand influences supply behavior. Externalities abound, and attention focuses on the incentives caused by the transmission of risk between firms. Prices, which reflect an average risk, do not provide the correct signals to firms about the effects of their actions on the other firms' costs.

In these uncertain markets, market operation, equilibrium, social welfare implications and incentives for vertical integration are drastically different from those of the classical markets. The traditional Pareto-Optimality implications of competitive markets disappear. Phenomena such as supply not equaling demand, shortages, concern with assured supplies, incentives for vertical integration, the risk of under or overproducing or of not fully utilizing the firm's capital stock -- all difficult, if not impossible to understand in the classical framework -- become easily understood when viewed in the more general framework of the models presented in the preceeding chapters. It is not true, however, that the models of this thesis represent a rejection of or alternatives to the classical supply and demand model. Instead, the models of this thesis represent more general models of market behavior, which include the classical model as a very special case.[1]

Consumer preferences between price and the probability of shortage determine equilibrium for the markets under study. Equilibrium is characterized by not only a price but also by a probability of satisfaction. In equilibrium, price will exceed the unit cost of production, and supply will not, in general, equal demand. By letting the size of the market increase, it was possible to show that the proportional risks of firm operation would decline and that percentagewise (for very large, perhaps unrealistically large, market size) the market equilibrium under uncertainty would approach that of the corresponding deterministic market. However, as market size increases, the absolute discrepancy between supply and demand in the equilibrium under uncertainty will, in general, grow arbitrarily large.

Although the differences between market clearing under uncertainty and certainty are evident, these differences become sharper when one considers

---

1 With the assumption of instantaneous production, the model of Chapter 2 is identical to the classical supply and demand model.

the social welfare implications of market operation. Unlike the traditional competitive markets, a competitive market under uncertainty does not necessarily lead to the socially preferred market operation. Usually, the government should use its taxing power to subsidize firms and, in this way, encourage firms to reduce price and expand production. Moreover, in increasingly uncertain times, competing firms will usually not provide responses that would stabilize the smooth functioning of the economy.

The heart of the analysis really centers on the transmission of uncertainty between firms. The stochastic nature of demand of one firm affects another firm's costs. The decision of firms to vertically integrate and produce some of the input for themselves affects the stochastic structure of demand in the input market. Price incentives are not sufficient to insure that firms will take account of the effect of their actions on the transmission of uncertainty to other firms.

Firms have an incentive to vertically integrate to obtain a more certain supply of inputs. Firms have an incentive not to integrate to avoid the risk of having unsold input. Strong incentives for a firm to vertically integrate arise because the vertically integrated firm is able to use its own inputs to satisfy its high probability demand and use the higher cost inputs purchased in the factor market to satisfy its low probability demand. Private incentives to vertically integrate are likely to exist, even though vertical integration can be socially undesirable. The result of such vertical integration is to lower the expected utility that a consumer can achieve, and usually to raise the price of the input and the final product, and to raise the probability that a consumer will be unable to satisfy his demands.

Competitive markets under uncertainty cannot be relied upon to properly allocate production and risk between interacting firms. However, prohibiting vertical integration solves one problem but creates another. It was found that the ability of an integrated firm to better coordinate the characteristics of its own internally produced input (i.e. price and probability of availability) makes it more likely that an integrated firm, and not a nonintegrated one, will have an incentive to develop and introduce new and beneficial technology.

The purpose of the models presented was not to model any one market in its full complexity, but rather to focus on the effects of demand and supply uncertainty on market behavior. There are many complications that could be added to the models to make them more realistic, but such complications would not alter the basic qualitative features of market operation. It was already argued in Chapter 2 why consumer search behavior and the holding of inventories would leave unchanged the main features of the analysis. For example, in Chapter 3, inventory holdings could be simply introduced into the model, and the only change in the analysis that need occur is that c and p would have to be replaced by c^* and p^*, where $c^* = c + h - \alpha c$ and $p^* = p + h - \alpha c$ where h = linear holding cost and α = discount rate.[1] The analytics change, not the basic insights into market operation. Another complication that might be added to the model would be to allow non-constant returns to scale. This complication would introduce the incentive for production smoothing. Still, regardless of which complications are introduced, the fact remains that it is the risk of not being able to fully utilize productive capacity, or to sell all the goods in stock, or to satisfy all potential customers that create incentives and produce results that the classical analysis, which ignores uncertainty, cannot comprehend.

Given that there are differences between market operation under certainty and under uncertainty, the questions arise as to how valid are

[1]See S. Karlin, "Optimal Inventory Policy for the Arrow-Harris-Marshak Dynamic Model"in K. Arrow, S. Karlin, H. Scarf, Studies in the Mathematical Theory of Inventory and Production, Stanford University Press, Stanford, California, 1958.

the assumptions of the models that have been presented and how significant are the incentives, especially those for vertical integration, that the models have found. Obviously, the answer to these questions depends on the particular market under study. Based on both everyday experience and on detailed descriptive studies of markets, it does appear that for many markets, the assumptions and the implications of the models under study are valid and useful in analyzing and understanding firm behavior. As the daily articles in the Wall Street Journal indicate, most businesses are very concerned about demand and supply uncertainty. For example, in a discussion of the reasons why a particular firm is so successful, we read that part of its success is due to its decision to supply some of its own foundry needs not only to cut costs but also to avoid the "continual" problems with uncertain deliveries from outside suppliers.[1] Such statements appear quite frequently and unless we are to disbelieve these statements, we must conclude that the classical analysis is missing some important aspect of market behavior. The fact is that for many markets there is no price mechanism that adjusts instantaneously to keep supply and demand always in balance. Prices do respond to market forces, and markets do clear -- but not necessarily in the classical fashion. Occasional shortages are a natural, not an unnatural feature of many markets.

Once it is evident that, for many markets, the assumptions of the models under study do appear reasonable, we must ask just how significant are the forces to vertically integrate that the models predict. Here too, the model seems to fit well with descriptive statements of industry studies.

[1] Wall Street Journal, June 4, 1975, p. 20.

In an in-depth study of the automobile industry, White[1] examines the reason why auto companies vertically integrate, and, most relevant for this discussion, provides a descriptive explanation of how risks motivate vertical integration. White's descriptive discussion echoes many of the points raised earlier.

In Chapter 5, we argued that vertical integration was a means of transferring risk between firms. White states ". . .integration is a two-edged sword. Though it reduces the risk of supply failure, it also converts variable costs into fixed costs - . . .More money is at stake, . . .the financial penalties of losses (that is, risks) have increased."[2]

We found in Chapter 5 that there would exist strong private incentives for vertical integration to occur, and identified the possibility for partial vertical integration. The strong incentives for vertical integration arise because the vertically integrated firm is able to satisfy its high probability demand by itself, and pass on the low probability demand to some other firm. In Chapter 5, we found that for the case of partial integration, the factor market acted as a type of insurance market for the final product firm, with the final product firm making less of a profit on any item that used a factor market input than on an item that used an internally supplied input. On these issues, White writes "A way of reducing the risks of vertical integration is through partial or tapered integration: a company can produce a portion of its needs of an item and buy the fluctuating remainder. This has the advantage of providing full utilization of its own equipment and allowing the

---

[1]L. White - The Automobile Industry Since 1945, Harvard University Press, Cambridge, Mass., 1971.

[2]*Ibid*., p.80.

suppliers to absorb the risk of fluctuations in demand. The company has to pay a premium to get someone else to absorb the risks, but the risk transfer is achieved. In the case of a supplier failure, production of the final good does not have to cease. . ."[1] "Tapered integration plays a large role in the industry."[2] When outside suppliers are used "Reliance on a single supplier has generally been avoided."[3] "More common is the practice of multiple suppliers. . ."[4]

Based on White's analysis and on the other descriptive studies cited in Chapter 5,[5] it does seem that the incentives identified in the models of Chapters 2 through 5, do indeed exert a significant influence on the decision of firms to vertically integrate.

This thesis has analyzed the behavior of markets characterized by demand and supply uncertainty, a non-instantaneously adjusting price, and a prior committment to production before demand can be observed. The study of the implications of the risks facing firms in this setting leads to a clearer and very different understanding of market operation than does the classical analysis which ignores the uncertainty in the marketplace. The models of the previous chapters are not alternatives to the classical model but instead are more general than the classical model and include it as a special case.

A key feature in understanding the operation of uncertain markets deals with the transmission of risk between firms, and these firms' response to

---

[1]White, op.cit., p.80.
[2]Ibid, p.83.
[3]Ibid, p.84.
[4]Ibid, p.85.
[5]See Section 5.1 for a literature review concerning the need for assured supplies creating an incentive for vertical integration.

their uncertain environment. Many characteristics of market behavior which are incomprehensible in the classical framework have a clear explanation when viewed in the more general framework of the models of the previous chapters. Supply not equaling demand, shortages, concern with obtaining assured supplied, the risk of under or overproducing or of not fully utilizing the firm's capital stock, all become natural features of market operation in the more general models. It is only by explicitly examining the effects of uncertainty on firms' responses that these features of market behavior together with their consequences can be fully comprehended.

BIBLIOGRAPHY

Areeda, P., "Structure Performance Assumptions in Recent Merger Cases" in S. Peltzman and J. Weston, eds., Government Policy Toward Mergers, Goodyear Publishing Co., 1968.

Arrow, K., "Vertical Integration and Communication", Institute for Mathematical Studies in the Social Sciences, Stanford University, 1974.

Arrow, K., "Vertical Integration and Communication", The Bell Journal of Economics, Spring, 1975.

Arrow, K., S. Karlin and H. Scarf, Studies in the Mathematical Theory of Inventory and Production, Stanford University Press, 1958.

Bork, R., "Vertical Integration and Competitive Processes" in S. Peltzman and J. Weston, eds., Government Policy Toward Mergers, Goodyear Publishing Co., 1968.

Brown, G., Jr. and M. Johnson, "Public Utility Pricing and Output Under Risk", American Economic Review, 1969.

Chandler, A., Jr., Strategy and Structure: Chapters in the History of of American Industrial Enterprise, M.I.T. Press, 1964.

Cichetti, C. J. and A. M. Freeman III, "Option Demand and Consumer Surplus: Further Comment", Quarterly Journal of Economics, 1971.

Coase, R., "The Nature of the Firm", Economica, 1937.

Fisher, F. M., "Stability and Competitive Equilibrium in Two Models of Search and Individual Price Adjustment", Journal of Economic Theory, 1973.

Fisher, F. M., "Quasi-Competitive Price Adjustment by Individual Firms: A Preliminary Paper", Journal of Economic Theory, 1970.

Green, J., "Vertical Integration and Assurance of Markets", Discussion Paper 383, Harvard Institute of Economic Research, 1974.

Hall, R. E., "An Aspect of the Economic Role of Unemployment" to appear in the Proceedings of the International Economics Association, April, 1975.

Iwai, K., "The Firm in Uncertain Markets and Its Price, Wage, and Employment Adjustments", Review of Economic Studies, April, 1974.

Karlin, S. and C. Carr, "Prices and Optimal Inventory Policy" in K. Arrow, S. Karlin and H. Scarf, Studies in Applied Probability and Management Science, Stanford University Press, 1962.

Kendall, M. and A. Stuart, The Advanced Theory of Statistics, Vol. 1, Hubner Publishing Co., 1969.

Malmgrem, H., "Information, Expectations, and the Theory of the Firm", Quarterly Journal of Economics, 1971.

Mc Call, J., "Probabilistic Microeconomics", Bell Journal of Economics and Management Science, 1971.

Mills, E., "Uncertainty and Price Theory", Quarterly Journal of Economics, 1959.

Mills, E., Prices, Output and Inventory Policy, Prentice-Hall, 1962.

Mueller, W., "Public Policies Toward Mergers" in S. Peltzman and J. Weston, eds., Government Policy Toward Mergers, Goodyear Publishing Co., 1968.

Peltzman, S. and J. Weston, eds., Government Policy Toward Mergers, Goodyear Publishing Co., 1968.

Rothschild, M., "Models of Market Organization with Imperfect Information: A Survey", Journal of Political Economy, 1973.

Rothschild, M. and J. Stiglitz, "Increasing Risk I, A Definition", Journal of Economic Theory, September, 1970.

Stigler, G., "The Division of Labor is Limited by the Size of the Market", Journal of Political Economy, 1951.

Visscher, M., "Welfare Maximizing Price and Output with Stochastic Demand: Comment", American Economic Review, 1971.

White, L., The Automobile Industry Since 1945, Harvard University Press, Cambridge, Mass, 1971.

Williamson, O., "The Vertical Integration of Production: Market Failure Considerations", American Economic Review, 1971.

Wilson, C., "A Model of Insurance Markets", mimeo.

Zabel, E., "A Dynamic Model of the Competitive Firm", International Economics Review, 1967.

BIOGRAPHICAL NOTE

Dennis W. Carlton was born February 15, 1951. He grew up in Brighton, Mass., and graduated in 1968 from Boston Latin School. He attended Harvard College, where he majored in Applied Math and Economics. He was elected to Phi Beta Kappa in 1971 and received the B.A. degree, Summa Cum Laude from Harvard in 1972. In September 1971, he married Jane R. Berkowitz. He received the Master's degree in Operations Research from M.I.T. in September 1974. Beginning in September 1975, Mr. Carlton will be an Instructor in the M.I.T. Department of Economics.