The Cost of Transacting

Harold Demsetz

The Quarterly Journal of Economics, Vol. 82, No. 1. (Feb., 1968), pp. 33-53.

Stable URL:

http://links.jstor.org/sici?sici=0033-5533%28196802%2982%3A1%3C33%3ATCOT%3E2.0.CO%3B2-Y

The Quarterly Journal of Economics is currently published by The MIT Press.

THE COST OF TRANSACTING *

Harold Demsetz

The empirical work in this paper is a study of the cost of transacting on a very important market, the New York Stock Exchange, but the economics of transacting, of which this paper is a beginning, has an importance that extends beyond particular markets and that argues against the general neglect accorded the subject by economists. This can be grasped by considering the operation of an economic system in which transaction cost is zero. The usual sources of inefficiency fail to exist in such an economic system.

Externalities? Persons subjected to harmful effects will, at zero negotiating and contracting cost, bring the value they attach to a reduction of these effects to the attention of whoever produces the harmful effects. The value placed on reducing these effects will be compared, again through costless negotiations, with the value attached by others to the benefits associated with the production of these effects. In response to such costless bidding by rivals, the output of harmful effects will be brought to efficient levels.

Monopolies? It will be profitable for agreements to be reached, again at zero transacting cost, between buyers and sellers that lead to efficient output rates in return for side payments. Efficient rates of output yield a bigger pie to be shared by all, so that in a milieu of costless transactions, utility maximization will yield efficient allocations even in the presence of monopoly and monopsony.

Of course, the existence of positive transacting cost has no direct relevance to economic inefficiencies. As with any cost, the question that is relevant for efficiency is whether or not the cost is appropriately economized. In some cases it will be efficient to have markets in which negotiations are carried forth to bring costs and benefits to bear on economic decision units. The value of realigning resources as a result of such negotiations is expected to be worth the cost of transacting. In other cases it will be efficient not to

negotiate; in a world of positive transacting cost some external and monopoly effects are consistent with efficiency. None of this makes unimportant the usual questions concerning the role of government in the resolution of externality and monopoly problems. For there are cases in which the cost of government action is less than the cost of transacting in markets. In such cases, we will employ government action that realigns resources more completely than can be achieved economically in the market place.[1]

In addition to these areas of application, the economics of transactions will be the core of an economic theory of money, for exchange will tend to be conducted in ways that economize on the cost of transacting.[2] An economic theory of money must inquire into what conditions favor the use of a specialized medium of exchange, for there seem to be circumstances in which barter appears more economical. "Portability, storability, and divisibility," words frequently marshaled to describe an economical money, become quantifiable characteristics when they are treated as methods for reducing the cost of transactions.[3]

The above subjects are beyond the scope of the present study, but the work presented below will shed some light on why the New York Stock Exchange (NYSE) has dominated for so long the organized trading of securities, and the empirical estimates to be given will enable us to reevaluate intuitive notions about imperfections in the capital markets; how much of the difference in borrowing costs between large and small firms can be attributed to differences in the cost of transacting rather than to imperfections in the capital market?

But the more general question we seek to investigate is the extent to which transaction costs are affected by the scale of trading. This aspect of trading, always of interest to economists, has implications that extend beyond the NYSE. Do standardization and delegation of transacting authority result in marketing scale economies? The NYSE provides us with an important source of data by which

1. A more detailed discussion of the relationship between transaction cost and efficiency can be found in R. H. Coase, "The Problem of Social Cost," *The Journal of Law and Economics*, III (Oct. 1960), and H. Demsetz, "The Exchange and Enforcement of Property Rights," *The Journal of Law and Economics*, VII (1964).

2. See A. Alchian and W. Allen, *University Economics* (Belmont, Calif.: Wadsworth Publishing Company, 1967), pp. 49–50.

3. The cost of using a specialized medium of exchange is essentially the cost of holding an inventory of money. The greater is the number of goods whose prices are relatively independent, the lower will be the risks associated with fluctuations in the value of this inventory, and the greater is the time rate of exchange, the lower will be the expected cost per transaction of inventorying money.

the presence or absence of scale economies can be detected, and what will be studied below in detail for this important market is an application of Adam Smith's famous proposition:

As it is the power of exchanging that gives occasion to the division of labour, so the extent of this division must always be limited by the extent of that power, or, in other words, by the extent of the market.

THE DEFINITION AND MEASUREMENT OF TRANSACTION COST ON THE NEW YORK STOCK EXCHANGE

Transaction cost may be defined as the cost of exchanging ownership titles. In the specific case of the NYSE, it is the cost of exchanging titles to money and to shares of stock. It is possible to increase or decrease this cost by a more or less inclusive definition of which activities are to be counted as transaction activities. From one viewpoint the cost of producing assets is necessary to the exchange of assets, whereas, from another viewpoint, only titles to assets need be produced for exchange to take place — the production of the assets themselves can be postponed indefinitely. And one could include in transaction cost the cost of being informed about the general nature of the market — the cost of making phone calls to one's broker or of reading the financial pages. Transaction cost is defined narrowly in this paper as the cost of using the NYSE to accomplish a quick exchange of stock for money. Broader interpretations lead to extremely difficult empirical and conceptual problems.

Given that titles to assets exist, given that decisions to exchange these titles have been made, and given that brokers or sales representatives have been informed of these decisions, what are the costs to buyers and sellers of using the NYSE to contract with each other? These remaining costs comprise transaction cost as the term is used in this paper. On the NYSE two elements comprise almost all of transaction cost — brokerage fees and ask-bid spreads. Transfer taxes could be included, but it is expedient to concentrate our attention on the two major components.

The inclusion of the ask-bid spread in transaction costs can be understood best by considering the neglected problem of "immediacy" in supply and demand analysis. Predictable immediacy is a rarity in human actions, and to approximate it requires that costs be borne by persons who specialize in standing ready and waiting to trade with the incoming orders of those who demand immediate servicing of their orders. The ask-bid spread is the markup that is

paid for predictable immediacy of exchange in organized markets; in other markets, it is the inventory markup of retailer or wholesaler.
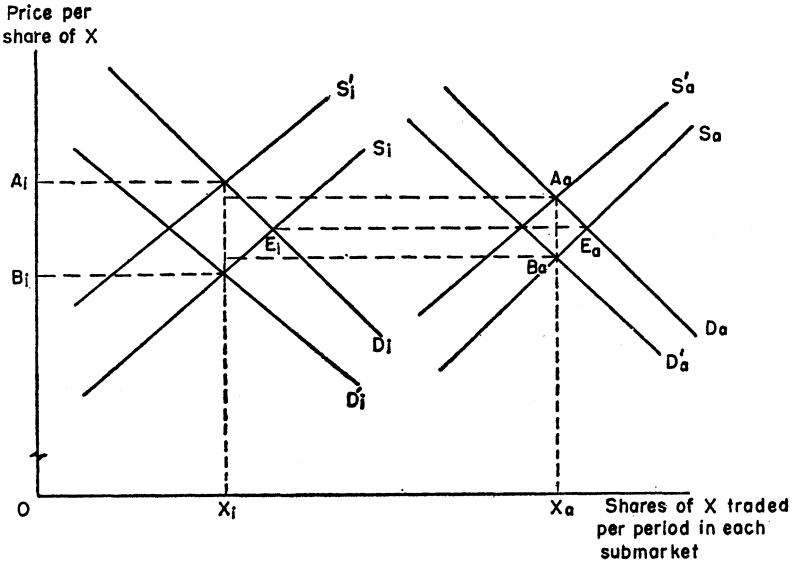

FIGURE I

In Figure I, curves D_i and S_i illustrate the demand and supply flows per unit of time for security X_i on the part of those who desire that their buy or sell orders be serviced immediately. Intersection E_i shows the conventional view of equilibrium price. Suppose that E_i is, in fact, the average price for which X has been and will be exchanged. But a person wishing to sell shares of X_i at price E_i cannot automatically count on the presence of a buyer. An order to buy X_i at price E_i may come to the market only after a time delay of minutes, hours, or even days. And a person wishing to buy X_i at price E_i has no guarantee that a sell order will be available with which to match his buy order. Thus, D_i and S_i do not illustrate *always present* market orders; rather, they measure time rates of demand and supply for which, at any given time, no market orders need be present.

Now let some person or persons provide the service of standing ready to sell or buy at stated prices immediately upon receipt of a matching order. To cover the cost of standing ready, these persons will be willing to buy X_i at a price that is slightly below E_i and sell at a price that is slightly above E_i. The difference between these two prices is the ask-bid spread. Their geometric counterparts can be illustrated in Figure I as follows: S_i is the supply curve of those

who wish to sell immediately. But persons willing to wait before selling, and who do so by keeping their selling offers active, must be compensated for the cost of waiting; to cover the cost of waiting, their supply curve will lie above S_i; S'_i shows the supply curve of asking prices for those who stand ready and waiting to sell to those who demand immediate servicing of their purchase orders. D_i is the demand curve for those who desire immediate purchases, so the intersection of D_i and S'_i yields the equilibrium ask price, A_i.

Similarly persons willing to supply securities according to schedule S_i require that these shares be sold as soon as they are offered. The service of immediately buying these shares is offered by persons willing to stand ready and waiting to buy as soon as sell orders reach the market. Waiting costs must be incurred by these persons, and, therefore, their demand prices will be somewhat lower than those who are not prepared to wait. A schedule of demand prices for those offering to wait for the arrival of selling orders is given by D'_i. The intersection of D'_i with S_i determines the equilibrium bid price, B_i, at which a sale can be executed immediately. The difference between A_i and B_i is the spread that is quoted on security X_i. There exist two equilibrium prices and not one — A_i for immediate sales and B_i for immediate purchases. E_i can be thought of as an arithmetic average of A_i and B_i. The spread measures the price of immediacy in both buying and selling securities and X_i (in the symmetrical presentation of Figure I) measures the volume of trading in each submarket. If buy orders and sell orders could be counted upon to arrive simultaneously, so that there would be no demand for the services of persons to stand ready and waiting, then E_i would determine the volume of trading. Without simultaneity, the volume of trading will be increased to the extent that those who stand ready and waiting specialize in reselling (rather than holding) what they have purchased. Thus, a specialist who buys at \$98 and sells at \$100 substitutes two transactions for what would be one transaction if the outside traders could count on their orders arriving simultaneously and at the same price, say \$99. (The curves with subscript a will be referred to later.)

A person who plays an important role in these submarkets on the NYSE is the specialist. The specialist earns his income in two ways: by managing orders and by assuming risk. The former role is to manage orders left with him by traders who desire to move to other positions on the floor of the exchange. In this role, the specialist acts as a broker; he matches buy and sell orders. If he matches an order left to his care with an order that is subsequently

presented to him by another floor trader, the specialist shares in the commission charged to the customer by the floor trader. This is the specialist's first source of income and in earning this income he serves as an information repository.

In his second role, the specialist may step in to match the order left with him by trading for his own account. If he does so, he acts as a trader and receives no part of the commission charged to the customer. Thus, if the first trader presents an order to sell (or buy) and the specialist buys (or sells) for his own account to match the trader's order, he does not earn any share of commissions on the exchange. However, such an operation can generate income for the specialist from other sources. He can engage in an opposite trading action at a preferential price differential later. If he buys for his own account, he can hope to resell later at a higher price than he paid; if he sells for his own account, he can expect to repurchase later at a lower price than he paid.

The specialist earns income through buying and selling for his own account by standing ready to step in during periods when ask-bid quotations submitted by outsiders are too far apart to keep trade active without wide jumps in price. The specialist can increase the rapidity of exchange with narrower price movements during such periods by offering a narrower ask-bid spread than outsiders are currently submitting. This role of the specialist involves judgment, investment, and risk-taking; it is a role that is difficult to computerize completely, although computer programs conceivably could aid the specialist in playing this role. The investment involved is common to that made by other inventory specialists such as retailers and wholesalers of commodities. It is the willingness to invest in inventory and to stand ready to exchange in order to offer quicker exchange at given cost to ultimate buyers and sellers. What makes the specialist important in this process is that he is obligated to fellow members of the exchange to make a market for the securities in which he specializes. If there exists no quotation from outsiders that is "reasonably" narrow, he must offer one of his own to facilitate trading. The specialist hopes, of course, to realize a profit on inventory turnover. Specialists in all types of markets perform essentially these same functions. All would like to acquire inventory at low prices and resell at high prices and to do so very rapidly, but competitive forces, to be discussed later, are at work in varying degrees in these markets and the stronger are these forces the closer will these markups be to the cost of waiting and carrying inventory.

It is apparent from the discussion that under competitive conditions the ask-bid spread, or markup will measure the cost of making transactions *without delay*. A person who has just purchased a security and who desires immediately to resell it will, on the average, be forced to suffer a markdown equal to the spread found in the market place. This markdown (plus brokerage commissions) measures the cost of an immediate round-trip exchange. Under less competitive conditions, this spread may somewhat exaggerate the underlying cost to those who stand ready and waiting of quick round-trip transactions, but, for any *given* degree of competition (since brokerage commissions do not vary with the time taken to complete a transaction), differences in spread will indicate differences in the cost of quick exchange. The typical spread for one security may be twice the percentage of price that it is for another; this can be taken to indicate that the cost of quick exchange per dollar invested in the first security is greater than it is for the second, and, perhaps, approximately twice as great. The spread, of course, can be thought of as measuring twice the cost of a one-way transaction; the last transaction price may be $40 and the currently quoted spread may ask $40½ and bid $39½, so that a market order pays a half point penalty relative to the last transaction price.

If the cost of quick exchange is higher for one asset than it is for another, we may assume that the cost of exchanging with any given time delay will be higher also, although not necessarily proportionately higher. The forces at work in determining the cost of quick exchange, we shall see, are not such that they can be expected to work in opposite directions if we increase the time interval during which an exchange is concluded. Hence, the analysis which follows can be expected to determine the identity of variables and to measure the direction of their effect on the cost of making transactions in highly organized markets whatever the time allowed to conclude an exchange. The *magnitude* of the effects measured, however, can be associated with *quick* exchange only.

The ask-bid spread and the commission brokerage are determined by different procedures and institutional arrangements. Generally, commission brokerage depends only on the price of a share and is independent of whether or not the executed order is a market or limit order. The relationship of commissions to prices is established collectively by members of the NYSE. The spread is determined by persons acting individually, by specialists, by floor-traders or by outsiders submitting market or limit orders. The

spread component of transaction cost will vary according to several aspects of the market for a security. The *structural* requirements for competition are more clearly in evidence in determining the spread than they are in determining brokerage commissions.

Data to be discussed later reveal that the spread comprises about 40 per cent and the commission brokerage about 60 per cent of the total transaction cost. Total transaction cost is approximately 1.3 per cent of the value of one round lot of a $48 stock.

THE DETERMINATION OF THE ASK-BID SPREAD

Persons who stand ready and waiting to transact with incoming market orders submit limit orders quoting a minimum ask price or a maximum bid price. The prices stipulated by incoming limit orders arriving "simultaneously" do not cross except by accident, perhaps because market conditions change between the time that a limit order is submitted and the time at which it arrives on the floor of the exchange. A prospective purchaser either submits a market buy order, in which case he concludes an exchange at the lowest existing limit ask price, or, if he is prepared to stand ready and waiting, he submits a limit bid order stipulating an upper limit to the purchase price that he is willing to pay. This upper limit will be below the currently effective ask quotation. Likewise, a prospective seller will never ask a price below the currently effective bid price. Transactions generally occur between a market order and a limit order, although sometimes market orders will be crossed if the quoted spread is relatively large and if a buy and sell order arrive simultaneously. Only by accident will two limit orders, a bid and an ask, be crossed in the market place.

Those who stand ready to wait for incoming market orders preempt early positions in the trading queue by offering relatively high bid quotations and relatively low ask quotations (although occasionally disparities in the size of orders will alter this priority). The lower the ask and the higher the bid offered by a trader, *ceteris paribus*, the shorter the period he expects to wait before concluding an exchange with incoming market orders. Quotations distant from the last transaction price, if they are maintained, produce transactions after a longer wait than quotations maintained closer to the last transaction price; price cannot skip over these closer quotations; they must be acted upon before price can deviate further from the last transaction price.

Any given set of limit orders and associated ask-bid prices will

be acted upon sooner the more rapidly market orders arrive, so that the expected waiting and inventory cost will be less for those who are at the front of the limit order queues. Of course, in response to lower waiting cost, the active securities may attract more limit orders and this would tend to delay the execution of those limit orders not at the head of the queue; but, in equilibrium, the average spread brought about for active securities by competing limit orders will be smaller than for inactive securities. Whoever is at the head of the trading queue will trade more quickly if the security is active; hence, to get to the front of the queue, traders will be willing to lower their asking prices and raise their bidding prices. The fundamental force working to reduce the spread is the time rate of transactions. The greater the frequency of transacting, the lower will be the cost of waiting in a trading queue of specified length, and, therefore, the lower will be the spreads that traders are willing to submit to preempt positions in the trading queue.

Waiting costs are relatively important costs for trading in organized markets, and would seem to dominate the determination of spread. An inverse relationship between spread and time rate of transactions is to be expected unless there are rapidly rising marginal costs of communicating with respect to the number of orders transmitted to a single market place. The technology of communicating prices and transferring titles in the security markets is such that offsetting diseconomies are likely to arise only on those occasions when trading is so heavy that queues of *market* orders are formed. Since these occasions, in fact, are few, it would seem that the organized trading of a security is subject to a *form* of scale economies. *This need not be true of all markets and more will be said later about this.* We assume here that *potential* waiting costs do, in fact, dominate security trading, and this leads us to the conclusion that trading is subject to scale economies on the NYSE, where scale refers to trading in a particular security and not necessarily to trading in a heterogeneous bundle of securities. Moreover, inventory and queuing theory leads us to believe that waiting costs will be reduced most rapidly when the transaction rate is small and increasing. The presence of economies in transacting is a major proposition to be tested in this paper.

The geometrical counterpart of our argument is illustrated in Figure I, where the spread generated by the curves with subscript i (for inactive security) is to be compared with the smaller spread generated by the curves labeled with the subscript a (for active security).

It is important to distinguish the relationship between waiting cost and *market* transaction rate in a particular security, which is what is described above, and that between waiting cost and the *individual trader's* transaction rate. A person who submits only a single limit order will expect lower waiting costs for an actively traded security than for an inactively traded security, as will a person who submits many limit orders. Nonetheless, for both active and inactive securities, it is possible for the individual trader's marginal cost of transacting to increase as the rate at which he transacts increases. As the volume of trading in a particular security increases, waiting cost curves *shift* downward for all market participants, both frequent and infrequent traders. Whether or not the marginal cost curve for waiting is negatively sloped is a separate question about which the evidence is conflicting.

Over recent years on the NYSE there has been a rapid attrition in the number of specialists who keep book on any single security. At the present time there is one specialist only for each security, although many specialists handle more than one security. Specialists are a minority of the membership of the exchange, and the other members have an interest in the maintenance of small spreads, since small spreads will increase the volume of business customers will do with them.

Since this increase in the specialization of specialists has been accepted and probably encouraged by the NYSE it would seem that members act as though they believe that scale economies are present with regard to the activities of a single trader. But if such scale economies do exist, it is strange that the percentage of trading by volume in which the specialist participates in his own behalf seldom exceeds 50 per cent and, when it is in excess of 50 per cent, it is usually for those securities that trade slowly. See Appendix I for supporting data.

The question is of some interest to us because scale economies with respect to the transactions of a particular trader suggest natural monopoly and the possibility of a divergence between the observed spread and the underlying cost of transacting. The question is not crucial since our interest here is in the fee paid by outsiders for transacting quickly on the NYSE and not necessarily on the underlying cost. Moreover, our primary concern is with the behavior of spread as the transaction rate increases and not with the absolute level of spread; even in the presence of some degree of monopoly power, the behavior of spreads in relation to transaction rates should give some indication of the behavior of the monopolist's underlying

cost. Nonetheless, the possibility of a divergence between spread and cost is important enough to warrant more attention.

Even though scale economies are present in the specialist's trading activities, there is little likelihood of his maintaining spread much above the cost of waiting. Competition of several types will keep the observed spread close to cost. The main types of competition emanate from (1) rivalry for the specialist's job, (2) competing markets, (3) outsiders who submit limit orders rather than market orders, (4) floor traders who may bypass the specialist by crossing buy and sell orders themselves, and (5) other specialists. Each of these competitive forces is considered briefly.

Competition affecting the spread will arise from rivalry for the job of specialist. It is in the interests of nonspecialist members of the exchange, who are a majority, to have small spreads maintained, for this will increase the time rate of transactions and the time rate at which commissions are earned. Indeed, one of the stated NYSE guidelines for specialists is that of making an orderly market by keeping spreads narrow and close to the last transaction price. If the spread is maintained according to the *conventional* profit-maximizing solution of natural monopoly and if this yields a large rent, it will be in the interest of nonspecialist members of the exchange to compete for the job of specialist by offering to maintain narrower spreads. Effective rivalry for the job of specialist should keep spreads close to the cost of providing a ready market for outsiders. Thus, although scale economies with respect to each trader's activity imply that each security will be handled by one specialist, it is not true that this necessarily means that the spread will be maintained at the monopoly levels indicated by a conventional natural monopoly solution to the specialist's maximizing problem.[4]

Many securities are listed on more than one exchange and trading (unlisted) securities is permitted on many more exchanges. The maintenance of a large spread on the NYSE is an invitation to trade the security elsewhere. The effectiveness of this competition is, of course, limited by the scale economies offered by the high transaction rate of the NYSE.

Outsiders who submit limit orders to the NYSE offer effective competition to specialists and manage to participate in large percentages of the transactions made on the exchange. An objection might be raised that outsiders must pay commission brokerage and, therefore, compete at a disadvantage with the specialist. But out-

4. The concept of rivalry introduced here and the difficulty it poses for the theory of natural monopoly are discussed more fully in my forthcoming paper, "Why Regulate Utilities," *Journal of Law and Economics* (Mar., 1968).

siders pay this brokerage whether they submit limit orders or market orders. Their choice between the two types of orders, therefore, should be unaffected by brokerage commission. One can say that higher brokerage, as with any other price, reduces the volume of transactions on the NYSE, but it cannot be said that it affects the probability that an incoming order will be a market order rather than a limit order. Moreover, part of the brokerage paid by customers is transferred to the specialist when he completes a transaction for others, i.e., when he acts as a broker; whereas, when he trades for his own account, he foregoes his share of a customer's commission. A relevant opportunity cost for the specialist when he trades for his own account is this foregone brokerage fee; if he is rational he will treat this as a cost of trading for his own account, just as an outsider should treat the brokerage he pays for trading on his account. Any financial barrier that must be overcome by outsiders in competing with specialists by submitting limit orders is probably slight.

Exchange members do not suffer from the limitations that might be assigned to outsiders. Members can compete with specialists by submitting limit orders of their own or by crossing orders for their customers, thus bypassing the specialist. If the spread quoted by the specialist is relatively large, exchange members can be expected to compete in this way.

Finally, specialists who handle different securities compete with each other. Offering a narrower spread than other specialists who handle different securities is one way a specialist can switch customers to his security and away from those handled by other specialists. Thus, a customer may ask his broker to secure quotations on several securities and allow differences in the spread (per dollar exchange) to affect his choice of which security to buy or sell. This is similar, for example, to the competition of furniture retailers who inventory highly similar furniture.

An enumeration of the forces of competition is not, by itself, convincing evidence of competition; the large share of trading in which the specialist does not participate is somewhat more convincing of the absence of "natural monopoly" conditions. Unfortunately, unless the underlying cost of waiting is known, it is not possible to assess the competitiveness of the spread. It may be possible, however, to get some idea of the strength of the competitive forces operating internally on the NYSE: A negative relationship between spread and the number of competing markets on which a security is listed can be interpreted as indicating that spreads would have been higher on some securities had trading on other markets

been prohibited. But, given the competition of other markets, the observed spreads cannot be taken as evidence of monopoly power since these other markets stand ready to compete and thus reduce the spread. The situation is complicated even further if we treat the listings on other markets as indications that the market in a particular security is more active than is indicated by the trading activity of the NYSE alone. There is some evidence that this is the case (to be discussed below), and, if it is, any negative relationship between spread and number of competing markets merely reflects the basic relationship between the spread and the transaction rate.

A security's price must also affect the spread quoted for quick exchange. Spread per share will tend to increase in proportion to an increase in the price per share so as to equalize the cost of transacting per dollar exchanged. Otherwise, those who submit limit orders will find it profitable to narrow spreads on those securities for which spread per dollar exchanged is larger. The tendency for these arbitrage activities to lead to spreads that are proportionate to prices will be strongest if the cost of transacting and of waiting is the same per dollar exchanged for high priced and low priced securities. Insofar as outsiders are concerned, the commission cost per dollar exchanged is lower for high priced securities. To the extent that this disproportionality reflects underlying cost differences per dollar exchanged, then members of the exchange also will face declining cost per dollar exchanged as security price is raised, and this should attenuate somewhat the tendency toward strict proportionality. The strength of this attenuation depends on how important is the lack of proportionality in brokerage commission. If the nonproportional aspects of brokerage commissions are small in relation to total exchange and waiting cost, only a minor and difficult to detect effect on proportionality will be present. The brokerage paid by members in trading for their own accounts is considerably less than that paid by outsiders. The risk of adverse price changes and the cost of waiting are relatively more important than commissions insofar as exchange members are concerned. At best, we can assert a strong positive relationship between spread and price and inquire into the sign of the second derivative.

The Determination of the Transaction Rate

The basic relationship that has been specified is between the spread and the transaction rate. The transaction rate may be higher or lower for a day or two, perhaps because of a short-lived

rumor, an accidental convergence of trading in the stock, or, more often, because the market for all stocks is temporarily active or inactive. It is not clear whether spreads will be related more closely to an average, normal, or long-run transaction rate or to the transaction rate of the moment. Two variables are used to measure the transaction rate: (1) number of transactions per day based on data for two (nonadjacent) days of trading and, (2) the number of shareholders. A few words about the use of number of shareholders is in order.

A plausible primary determinant of a long-run or "normal" transaction rate for assets belonging to a given asset class (shares of stock in different companies, automobiles of different makes, etc.) is the number of asset owners. A doubling of the number of participants will approximately double the transaction rate. Unless there is underway a sudden change in demand for the security, the number of persons presently owning shares will be positively related to the number of market participants interested in bidding on the asset. We expect, then, that the number of shareholders will be strongly related to long-run transaction rates.

There is a minor statistical advantage to number of shareholders. The statistical work to follow is based on a random sample of 200 securities traded on the NYSE. Observations are recorded for two days (separated by about a month). Six securities were not traded both days and these cannot be used directly in regressions incorporating logarithmic or inverted forms of the transaction rate. No such problem is encountered with the number of shareholders.

Statistical Results

The following symbols identify the variables discussed above.

S = the ask-bid spread measured in dollars per share.

T = the time rate of transactions defined as the number of separately recorded transactions per day on the board of the NYSE. A single transaction may be one in which any number of shares is traded provided only that it is reported as a single transaction. A trade of 500 shares may represent an exchange between two persons, which it ordinarily will during the course of a day's trading, or more than two persons, which it may represent when the market opens and is cleared of overnight orders.[5]

---

5. The number of transactions and the volume traded per day are very highly correlated.

P = the price per share.

N = the number of shareholders measured in hundreds.

M = the number of markets on which the security is listed.

The data used in the statistical analysis are a random selection of 200 securities listed on the NYSE. Incomplete data for eight of these reduced the sample size to 192. Data on spread, price, and transaction frequency are calculated from the *Francis Emory Fitch Sheets*.[6] Observations on these variables are averaged for trading on two days, January 5 and February 28, 1965. The number of shareholders for each security is taken from *Moody's* and the number of markets from Standard & Poor's *Stock Guide*. Essentially, then, the data consist of cross-sectional measures of the variables.

Our interest is primarily in obtaining regression estimates for equations (IA) and (IB) and, secondarily, for equation (II).

(IA) $S = f(T, P, M)$

(IB) $S = g(N, P, M)$

(II) $T = h(N)$

Indicating partial differentiation by subscripts, we expect to find the following constraints operative for equations (IA) and (IB).

(1) $S_T < 0$, (2) $S_{TT} > 0$, (3) $S_p > 0$,

(4) $S_N < 0$, (5) $S_{NN} > 0$.

There is ambiguity about the sign of T_{NN} and some reason to believe that (6) $S_{PP} \le 0$ and (7) $S_M \le 0$. Equation (II) will be discussed later.

Different transaction rates can be pictured geometrically as in Figure I. Smaller transaction rates are represented by the supply and demand curves located closer to the vertical axis and larger transaction rates by supply and demand curves farther to the right. The long-run transaction rate is determined primarily by exogenous forces, of which the dominating force seems to be N (for example, N turns out to give a slightly better fit than number of shares outstanding). The price of the security also is fundamentally determined by exogenous forces such as anticipated corporate earnings. The role of M, the number of markets in which the security is listed, is not obvious. It is inserted to gauge the degree to which listing on outside markets lowers the spread. But it must be remembered that M may also be related to trading activity, so that, for any given value of N or T, securities listed on more exchanges may tend to be more actively traded. Both views of the role of M suggest that if increases in M play a role, it will be to reduce S.

6. Francis Emory Fitch, Inc., 138 Pearl Street, New York, N.Y.

It is important to note that it is not our purpose to estimate any of the curves shown in Figure I. We seek an explanation for any difference in spread between security i and security a. Our discussion suggests that prices and (short- or long-run) transaction rates will dominate as forces affecting the spread. The fee paid to others for standing ready and waiting to transact immediately is the dependent variable. If this fee is approximately competitive, it measures the cost per share of standing ready and waiting; this cost function is, we suppose, invariant with respect to the identity of the share traded, so that, in the competitive case, spreads are viewed as points on this stable cost function.

If the fee for trading quickly is not approximately competitive, difficulties arise in any attempt to estimate the underlying cost function; the margin above cost reflected by spreads in the absence of competition need not be the same for different securities. In the noncompetitive case this margin will be affected by the shape of the underlying marginal cost curve for waiting and by differences in the elasticity of demand for quick exchange. Coping with these problems requires a model in which variables are included that are expected to be related to differences among securities in these respects. No attempt is made here to view the market as noncompetitive.[7] But even though the spread be determined in noncompetitive markets, a negative association between spread and transaction rate will indicate that waiting cost per share is negatively related to transaction rates (unless the demand for transacting quickly becomes less elastic as the transaction rate increases).

Least-square estimates of equations IA and IB have been calculated using both linear and logarithmic forms. The estimates for these are presented in Appendix II. The best-fitting regressions are linear in P and logarithmic in T and N. Linearity in P gives *slightly* better results than the use of $ln P$; however, the coefficient of $ln P$ is highly significant and conforms to the condition $S_{PP} < 0$. $ln T$ and $ln N$, however, seem to give much better fits than T and N. The best-fitting regressions are presented below. The t statistic is given in parentheses and the square of the correlation coefficient and the standard error of estimate are also indicated. For the sample size employed here and assuming conditions of normality, t values above

---

7. Assuming that the spread is determined in markets best approximated by the competitive model is *not* the same as assuming that commission brokerage to outsiders is set competitively. Whatever is true about how the membership of the NYSE decides collectively (with the cooperation of the SEC) about brokerage commissions, it is not true that the spread is set collectively; nor is it true that the interest of most members of the NYSE is served by spreads that are larger than underlying costs.

2 indicate significance levels above .95 probability for one-tailed tests of significance. Residuals from the equations are very well behaved.

(IA) [8] $S = .38027 + .0080709\,P - .11527\,ln\,T - .022906\,M$
  (5.80)   (12.15)   (6.46)   (.97)
  $r^2 = .57637$ $s_{y.x} = .28632$

(IB) $S = .48411 + .0088554\,P - .080911\,ln\,N - .029691\,M$
  (5.65)   (13.24)   (5.10)   (1.21)
  $r^2 = .5422$ $s_{y.x} = .29666$

Both regressions give highly similar fits; although (IA) gives a slightly better fit, the use of number of shareholders does surprisingly well. All coefficients take on the expected algebraic signs and all except the M coefficient are highly significant. The coefficients of $ln\,T$ and $ln\,N$ yield the expected second derivatives. The coefficient of M cannot be judged to differ significantly from zero in the light of the evidence presented here. The reader will note that the significance of the M coefficient increases slightly when $ln\,N$ is used in place of $ln\,T$. The reason for this is that M is associated slightly with differences in transaction rates that are not explained by differences in N. This can be seen below where regression estimates of two versions of equation (II) are given.

(II) $T = 9.4897 + .02263\,N$
  (6.72)   (14.49)
  $r^2 = .52489$ $s_{y.x} = 18.69261$

(II′) $T = 10.837 + .03744\,P - 11.811\,S$
  (2.89)   (.65)   (2.72)
  $+ 1.6527\,M + .02166\,N$
  (1.09)   (13.70)
  $r^2 = .55667$ $s_{y.x} = 18.20096.$

All coefficients take on signs that are consistent with the forces operating in the market. N dominates in both (II) and (II′), so that T is determined primarily by exogenous forces. Note that r^2 and $s_{y.x}$ do not improve significantly when P, S, and M are added. The signs of the coefficients of P and S are consistent with the roles played by these variables. Higher values of P, given S, imply lower transaction costs per dollar exchanged and, therefore, higher transaction rates. And higher values of S, given P, imply higher transaction cost per dollar exchanged and, therefore, lower transaction rates. The dominance of N undoubtedly reflects a strong tendency

8. Regression estimates in (IA) contain six fewer sample observations than in (IB) because $T = 0$ for six securities.

for S and P to be adjusted to their equilibrium values, so that the cost of transacting per dollar exchanged does not vary much from one security to another; N, however, does differ between securities and, given the equilibrium cost of transacting per dollar, accounts for most of the differences in observed transaction rates.

The (not significantly) positive coefficient of M in (II') is consistent with the belief that, for given N, differences in M will be associated positively with differences in T. This is consistent with the (not significant) negative relationship between S and M indicated in equations (IA) and (IB). Finally, it should be noted that results not presented here reveal that the use of $ln\ N$ significantly impairs the fit of regressions (II) and (II').

SUMMARY AND COMMENTS

The statistical analysis strongly indicates that the cost of exchanging a security declines as trading activity in that security increases. It would seem that the centralization of trading activity on the NYSE can be explained by the lowering of transaction costs thereby achieved. These results probably hold for organized trading in commodities and currencies also. Adding additional securities or different commodities may increase transaction costs and eventually limit the number of securities or assets traded in a given market. Should this be the case, and this certainly has not been established here, we would have a plausible explanation for the dominance of the NYSE in trading the securities it does list and, at the same time, an explanation for the NYSE's self-imposed curb on the number of corporations listed.

The distinguishing characteristic of such trading on organized exchanges is the willingness of customers to forego a personal examination of the goods bought and sold. This allows a high degree of standardization and enables communicating and title exchange costs to be kept low even for large transaction rates. When customers are willing to let others buy and sell for them, when they are willing to conclude an exchange without a personal prior examination of the goods, the concentration of trading on relatively few markets offers the prospect of lower transaction costs.

When customers are unwilling to forego a personal examination of goods, markets must be able to cope with many more bodies, both human and automobile, per transaction. The associated crowding and the need for product display (as is required when shares of different corporations are traded) probably generates diseconomies.

However, as long as it is possible to multiply the number of markets without significant increases in the rent earned by factor inputs, scale diseconomies for single markets need not and do not seem to produce scale diseconomies for the marketing industry as the number of competing markets grows.

In concluding this paper, it is interesting to examine the implications of the statistical evidence for measuring imperfections in capital markets. Differences in borrowing rates often are taken as indications of such imperfections, but differences in borrowing rates cannot be taken at face value since cost differences may exist that account for a part or all of the observed differences in rates. The cost of transacting is relevant to this problem since our evidence suggests that the markups required to market additional shares will be less for larger companies than for smaller companies if we accept what most certainly is true, that larger companies tend to have more shareholders.

Assume that two corporations desire to increase outstanding equity by the same percentage of existing equity. One corporation already has (or anticipates that it will have) 100,000 shareholders ($N = 1000$); the other has 1000 shareholders ($N = 10$). Assume further that both shares will sell at *roughly* the same price so that commission brokerage per round lot is identical. The cost of inventorying the securities and of waiting will differ for the two issues according to the evidence presented above, so that we should expect the larger corporation to secure a price per share somewhat higher than the smaller corporation. The difference in price per share will reflect the difference in spread, or distribution markup, per share. This can be estimated from equation (IB). Holding price and number of markets constant, the difference in numbers of shareholders produces a difference in spread equal to about 37c per share. If the prices per share are assumed to be in the neighborhood of $40, then for every $40 secured by the smaller company $40.37 will be secured by the larger company, a difference equal to almost 1 per cent in borrowing cost.

The results are only slightly less impressive if transaction cost per dollar exchanged is used directly as the dependent variable. Define: $X = S/P$. Using least-squares to regress X on N and M yields

(III) $X = .019224 - .0015284 \ln N - .00021483 \, M.$
$$(11.58) \qquad (4.44) \qquad\qquad (.40)$$

A difference in size of corporation of the magnitude indicated in the previous example generates a difference in borrowing cost equal to

.7 per cent. If the larger corporation has one million shareholders, the difference in borrowing cost becomes 1 per cent. Of course, these calculations are based on the presumption that differences in transaction cost will be passed on to the borrower.

APPENDIX I

Table I is adapted from the report of the *Special Study of Security Markets*. The tabulation contains the percentage of stock days (defined as one day's observation on one stock) falling into each cell. The cells relate the percentage of market transactions in which the specialist participated (the columns) to market volume. For example, the entry in row 1, column 1 indicates that of those stocks trading 10,000 and more shares a day, 42.5 per cent were observed to fall into the specialist participation rate category of .01 per cent to 16.00 per cent of market volume. Large cell entries are underlined twice and small cell entries are underlined once. It is clear that the more inactive the security, the greater the specialist participation rate tends to be.

TABLE I *

NEW YORK STOCK EXCHANGE MARKET VOLUME VS.
SPECIALIST TRANSACTION PARTICIPATION RATE
(Per cent of stock days ᵃ in each category of market volume falling
into each range of specialist transaction participation rate)

Specialist Transaction Participation Rate ᵇ

Market Volume (Shares)	.01% to 16.00%	16.01% to 26.00%	26.01% to 37.00%	37.01% to 50.00%	50.01% and over	Total
10,000 and over	42.5%	18.9%	18.9%	8.0%	1.7%	100.0%
5,701 to 10,000	28.8	26.7	24.8	15.0	4.7	100.0
3,901 to 5,700	25.9	23.1	25.7	15.6	9.7	100.0
2,801 to 3,900	25.8	19.9	23.8	18.9	11.6	100.0
2,101 to 2,800	24.7	19.1	23.7	18.1	14.4	100.0
1,501 to 2,100	22.5	19.4	20.7	18.5	18.9	100.0
1,101 to 1,500	17.7	19.1	22.4	18.8	22.0	100.0
701 to 1,100	14.4	18.7	19.3	21.5	26.1	100.0
401 to 700	4.1	26.2	13.7	25.1	30.9	100.0
10 to 400	0.4	7.9	15.2	25.1	51.4	100.0

* *Source*: Adapted from *Report of the Special Study of Security Markets*, 88th Congress, House Doc. No. 95, Part 2, p. 459.
a. One stock day = a one day observation on one stock. Data are for three weeks ending January 27, March 24, and June 16, 1961.
b. Number of specialist transactions as a per cent of the number of market transactions.

Appendix II

Linear forms:

(A) $S = .2101 + .008996\,P - .003589\,T - .04712\,M$
     (3.45)   (13.22)   (4.42)   (1.91)
       $r^2 = .52689$ $s_{y.x} = .30035$

(B) $S = .1787 + .009275\,P - .00008871\,N - .05477\,M$
     (2.90)   (13.30)   (3.44)   (2.17)
       $r^2 = .50857$ $s_{y.x} = .30611$

Log form:

     $S = -.2407 + .3522\,ln\,P - .1118\,ln\,N - .02808\,M$
     (1.56)   (10.37)   (6.41)   (1.05)
       $r^2 = .4617$ $s_{y.x} = .32036$

Best fitting forms (used in text):

(A) $S = .38027 + .0080709\,P - .11527\,ln\,T - .022906\,M$
     (5.80)   (12.15)   (6.46)   (.97)
       $r^2 = .57637$ $s_{y.x} = .28632$

(B) $S = .48411 + .0088554\,P - .080911\,ln\,N - .029691\,M$
     (5.63)   (13.24)   (5.10)   (1.21)
       $r^2 = .5422$ $s_{y.x} = .29666$

UNIVERSITY OF CHICAGO