Discussion Paper No. 570

# Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders

by

Lawrence R. Glosten*
Northwestern University

and

Paul R. Milgrom
Yale University

August 1983

Abstract

We examine properties of bid and ask prices quoted by a risk-rental, zero-profit specialist facing heterogeneously informed traders. The presence of traders with superior information leads to a positive bid-ask spread. The resulting transaction prices converg information and if volume of trade (n) increases the spread decreases with the average squared spread in the order of 1/n. An increase in insider activity or an increase in the quality of their information leads to larger spreads. A bid-ask spread implies a divergence between observed returns and realizable returns. Observed returns are approximately realizable returns plus what uninformed anticipate losing to informed.

*Lawrence R. Glosten
J.L. Kellogg Graduate School of Managmeent
Northwestern University
Evanston, Illinois 60201

## Acknowledgements

Milgrom's research was supported by National Science Foundation Grant IST - 8208600. We wish to thank Rolf Banz (INSEAD), Dan Galai (UCLA), Jay Ritter (Wharton) and the participants in the University of Chicago Finance Workshop for helpful comments.

## I. Introduction

The usual view of markets is as a place where buyers and sellers come together and trade at a common price, the price at which supply equals demand. Securities exchanges are often pointed out as excellent examples of markets that operate this way. In fact, however, trading on exchanges takes place over time, and some institutional arrangements are necessary to help match buyers and sellers whose orders arrive at different points in time. On exchanges like the New York Stock Exchange, the economic function of the specialists and the floor traders is to match traders in this environment where supply and demand cannot be equal at each point in time.

The matching problem is most acute in trading shares of small companies, where the volume in trade is relatively low. A common problem in this environment involves the number of insiders who trade in the shares relative to the total trading volume. Many questions arise; among them: How completely do prices reflect insider information (Fama (1970))? How large are insider profits? How does the specialist behave in this environment?

A number of researchers have examined the optimal behavior of a specialist and how it leads to a bid-ask spread. The usual approach examines the management of inventory by a monopolist specialist, concentrating on the effect that inventory costs have on the bid ask spread; e.g. Ho and Stoll (1981) Amihud and Mendelson (1980) and Garman (1976).

The approach pursued here, is to show that the bid ask spread can be a function of informational differences, even when there are no inventory costs and competition forces the specialist's profit to zero. Specifically, the specialist faces an adverse selection problem, since a customer agreeing to trade at the specialist's ask or bid price may be trading because he knows something that the specialist does not. In effect, then, the specialist must offset the losses suffered in trades with the well informed by gains in trades with liquidity traders. These gains are achieved by setting a spread. This informational source of the spread has also been suggested by Bagehot (1971) and formally analyzed by Copeland and Galai (1980).[1]

In this paper, we posit a risk neutral competitive specialist. The specialist may have to compete for orders with floor traders, or there may be competition among specialists at geographically separated markets. This specification allows the derivation of the transaction price process and an investigation of its informational characteristics. We find in Section II that the transaction price process exhibits a semi-strong form of efficiency, and if per period trading is allowed to increase indefinitely, that both the spread and the gains to superior information disappear. We also examine the spread and find that it is larger when, other things equal, the proportion of insiders is greater or their information is better or the liquidity trading motives of uninformed investors is weaker.

In section III, we examine a variation of the market model in which a fixed postive return is expected by uninformed traders. The existence of a spread implies that using the transaction price process to measure holding period return overstates the realizable returns. Further, we show that such observed returns are approximately a normal return plus a return available only to insiders. This along with the results of section II identifies one possible source of such anomalies as the small firm effect (Banz (1981) ) and the ignored firm effect (Arbel and Strebel (1981)). The evidence suggests that much of the excess returns on small firm investments occur in January. If the annual report for small firms tends to contain a considerable amount of new information and if insiders have early access to that information, then our analysis would predict an especially large spread in the period before the report is made public and presumably after the end of the firm's fiscal year. This combined with a "required rate of return" assumption will lead to observed large returns and normal realizable returns.[2]

Section IV indicates areas of future research and provides concluding comments.

## II. The Basic Model

The market that we are modelling is a pure dealership market, i.e., the specialist performs no brokerage services, and in effect all orders are market orders. Trade occurs according to the following sequence of events. The specialist sets a bid and ask price with the interpretation that he is willing to sell one unit of stock at the ask and buy one unit of stock at the bid. An investor arrives at the market and is informed of the bid and ask at which time he is free to buy one unit at the ask or sell one unit at the bid or leave. The specialist is free to (and in general will) change the bid and ask at any time after an arriving investor has made a decision and before the next arrival of an investor. That is, if an arriving order leads to a trade, the trade takes place at the quoted bid or ask. After the trade, the specialist may revise the bid and ask.

To examine the informational characteristics of such a market, we assume that there are informed investors and purely "liquidity" traders. At some time $T_0$ in the future, some random dollar value $V$ ( $V \geq 0$, $Var(V) < \infty$ ) per share will be realized, and the informed have information about this random variable $V$. Time $T_0$ may be interpreted as the time at which some informational event occurs — an earnings announcement, for example. At that time, there will be agreement on the value of the firm and the informational differences between insiders and outsiders will be minimal. The informed receive information sequentially and their orders are placed sequentially. When an informed trader comes to the market, his decision to buy, sell or leave is based on his information. The informed trader may be speculating based on inside information or superior analysis, or he may simply have a "liquidity" reason for trading. We will refer to the informed traders as insiders though other interpretations are possible.

As is typical (and necessary) in models that have traders with superior information, the active participation of uninformed is posited exogenously. Thus, we assume that there are uninformed investors, each of whom is willing to buy ( or sell ) if the price is below ( or above ) his personal reservation price. We assume that investors arrive one by one, randomly and anonymously at the specialist's post; these arrivals form an arbitrary arrival process.[3]

All participants, informed, uninformed and the specialist, are risk neutral. Each participant assigns random utility to shares of stock, x, and current consumption, c, as $u(x,c|\rho) = \rho x V + c$, where $\rho$ is a parameter of the individual investor's utility function representing his personal trade-off between current and future consumption derived from ownership of the asset. For the specialist, we take $\rho = 1$; this is just a normalization. Generally, a high $\rho$ indicates a desire to invest for the future; a low $\rho$ indicates a desire for current consumption. This "liquidity parameter" could be the result of imperfect access to capital markets or it could represent differential subjective assessments of the distribution of the random variable V. The risk neutrality assumption implies that in order for there to be trade, there must be some variation in $\rho$ across market participants, for otherwise the "no trade theorem" of Milgrom and Stokey (1982) will imply that the spread is large enough to preclude all market activity. Since $\rho$ is to be unknown to the specialist, and a pure preference parameter, we treat it as a random variable independent of V and any information about V and independent across traders. We allow the possibility that $\rho$ might follow a different distribution for the informed and the uninformed.

Investors, upon arriving at the market and hearing the bid and ask, maximize expected utility given their information to date. For uninformed investors, this information consists of all past transaction prices, the current bid and ask as well as any publicly available information. The informed also have access to the previous transaction price sequence, the current bid and ask, and all public announcements, but in addition they have been able to see some private signal. Formally, let $H_t$ denote the information available publicly, up to clock time t. If an uninformed investor arrives at time t, then his information, upon arrival, is $H_t$ joined with the information generated by the quoted bid and ask. If an arrival at time t is informed, then his information is represented by the join of his private information $I_t$, the information generated by the quoted bid and ask and the public information $H_t$. By including the specification of who is informed in the sample space, we can generally represent the information of an arrival at time t by $F_t$, a refinement of $H_t$, including the information conveyed by the quoted bid and ask.

Putting the utility functions and information structures together, the optimal decision of an investor arriving at time t, given bid B and ask A is given by

buy if $Z_t > A$

sell if $Z_t < B$

where $Z_t$ is given by

$Z_t = \rho_t E[V|F_t] = \rho_t(1-U_t)E[V|H_t \vee I_t,A,B] + \rho_t U_t E[V|H_t,A,B]$, where $U_t$ is one if the arrival at t is uninformed and zero otherwise.

Given the above behavior of the market participants, the specialist chooses bid and ask prices. Let the information available to the specialist at time t be represented by $S_t$. Assuming anonymity, the specialist cannot know when the bid and ask prices are set whether the next customer will be an insider or an outsider. Given the investors' behavior, the specialist's expected profit from an arrival at time t, given the information available to him at time t, $S_t$, and bid and ask prices B and A is:

$$E[(A-V)I_{\{Z_t > A\}} + (V-B)I_{\{Z_t < B\}} \mid S_t]$$

or

$$(A - E[V|S_t, Z_t > A])P\{Z_t > A|S_t\} - (B - E[V|S_t, Z_t < B])P\{Z_t < B|S_t\}.$$

The above holds as long as there are zero costs associated with all short positions in cash or stock. Our central assumption about the specialist is that he earns zero expected profits on each purchase and each sale. To illustrate how competition might lead to such a description, suppose there are two specialists in this one stock. Both have the same information and face the same population. Suppose the first specialist sets an ask price $A^1$ so that $A^1 > E[V|S_t, Z_t > A^1]$. The second specialist will rationally undercut the first by choosing an ask, $A^2 < A^1$ and $A^2 \geq E[V|S_t, Z_t > A^2]$. The competitive, (or Bertrand) equilibrium at time t (if it exists) is a pair of functions $A_t$ and $B_t$ satisfying

$$A_t(\omega) = E[V|S_t, Z_t > A_t(\omega)](\omega),$$

$$B_t(\omega) = E[V|S_t, Z_t < B_t(\omega)](\omega).$$

Where $Z_t = \rho_t E[V|F_t]$ and $A_t$ and $B_t$ are measurable with respect to $F_t$ (i.e., the customer knows the bid and ask prices).

General existence of such functions would be difficult to show, since it involves a "rational expectations" type of fixed point condition. The definition is not vacuous, however, as the following examples show. If the specialist's information, $S_t$, is a finer partition than the information of the informed, then $A_t$ and $B_t$ will both be equal to the conditional mean of V given the information $S_t$. If, on the other hand, the specialist's information is the same as the publicly available information $H_t$, then $A_t$ and $B_t$ are given by:

$$A_t = \inf\{a\colon a \geq E[V|H_t, Z_t > a]\}$$

$$B_t = \sup\{b\colon b \leq E[V|H_t, Z_t < b]\}$$

To insure that the customer's decision rule is formally well defined, and to illustrate the source of the spread, we must prove that at all times the ask exceeds the bid and, if insider trading actually occurs--or more precisely if it could occur--that the expectation of V lies strictly between the bid and ask. This proof and a later one both rely on the facts from probability theory that for any random variable X with finite expectation, $E[X|X > a] \geq E[X]$ and $E[X|X \geq a]$ is nondecreasing in a. The inequality is strict whenever $0 < P\{X > a\} < 1$.

Henceforth, we shall use $E_t$ to denote conditional expectations given the common knowledge at time t, i.e., $E_t[\ .\ ] = E[\ .\ |\ S_t \quad F_t]$, where the "meet" $S_t \quad F_t$ denotes the events which are in both $S_t$ and $F_t$. Notice that $A_t = E_t[A_t]$ and $B_t = E_t[B_t]$, since the bid and ask prices are always common knowledge and hence are effectively constants at time t. Also, our informal assumptions about $\rho_t$ (that it conveys no information about V nor about an informed trader's opinions) can be adequately formalized by:

(i) $E_t[V|F_t, \rho_t] = E_t[V|F_t]$ and

(ii) $E_t[E_t[V|F_t]|\rho_t] = E_t[V]$.

<u>Proposition 1</u>

Suppose equilibrium bid and ask prices exist satisfying the zero expected profit conditions:

$$A_t = E[V|S_t, Z_t > A_t]$$

$$B_t = E[V|S_t, Z_t < B_t].$$

Then the ask price is greater and the bid price is less than the expectation of V: $A_t \geqslant E_t[V] \geqslant B_t$. The inequalities are strict if adverse selection is possible, i.e., if

$$P\{Z_t > E_t[V],\ E_t[V|F_t] > E_t[V]\} > 0 \text{ and}$$

$$P\{Z_t < E_t[V],\ E_t[V|F_t] < E_t[V]\} > 0.$$

Proof

We prove only the first inequality, since the proof of the second is similar. Also, for brevity, we omit the time subscripts. Let C be the event that the customer makes a purchase:

$C = \{Z > A\} = \{E[V|F] > A/\rho\}$. Then, $A = E[V|S,C]$ so

$$A = E[A|C] = E[E[V|S,C]|C] = E[V|C] = E[E[V|C,\rho]|C]$$

$$= E[E[E[V|F,C,\rho]|C,\rho]|C] = E[E[E[V|F,\rho]|C,\rho]|C]$$

$$= E[E[E[V|F]|C,\rho]|C] \geqslant E[E[E[V|F]|\rho]|C]$$

$$= E[E[V]|C] = E[V].$$

If the additional condition stated in the proposition holds, then the inequality is strict. Q.E.D.

The proof of the above proposition shows why the competitive specialist must maintain a spread. If the ask is not set higher than the conditional expectation, $E_t[V]$, then the specialist will lose money on average on transactions with informed traders. He will not make money on transactions with uninformed traders, for as soon as an uninformed trader sees such an ask he will have as much of the specialist's information as he needs. Thus, the specialist, setting such an ask price will lose on average. If the proportion of liquidity traders is high or if their transactions motives are sufficiently strong or if the insiders' informational advantage is small, then there will exist bid and ask prices at which the losses to informed traders will on average balance the gains from uninformed traders. These prices involve a spread that is due solely to the adverse selection phenomenon and not to any risk aversion or monopoly power of the specialist. If these conditions fail to hold, the equilibrium ask may be so high and the bid so low that no trade can take place. When this happens, we say that the market has shut down. The market will "reopen" when enough public information emerges to neutralize the insiders' advantage.

We now assume that both the specialist and the traders see the results of every trade, and know the quoted bid and ask. Let $H_t^+$ and $S_t^+$ be respectively the information available to the uninformed and the specialist just after a trade at time t. These fields include information about whether a trader has arrived at time t, whether he bought or sold, and the price at which trade occured.

Let $T_k$ be the times at which trades occur. The above discussion shows that the $T_k$ are stopping times relative to $\{S_t^+\}$ and $\{H_t^+\}$. Define $S_k$ and $H_k$ by $S_k = S_{T_k}^+$ and $H_k = H_{T_k}^+$. (Also, any process subscripted with a k will be understood to be the value of the process at time $T_k$). If the kth trade takes place at the ask at time t, i.e., there is an arrival at time t and $Z_t$ exceeds $A_t$, then the transaction price will be the ask price, which in this event is equal to $E[V|S_k]$. Similarly, if there is a trade at the bid, the transaction price is the bid price which in this case is also given by $E[V|S_k]$. This observation allows us to write the kth transaction price as $P_k = E[V|S_k]$, as long as k trades take place (i.e., as long as $T_k$ is finite). This motivates our definition of the transaction price process as $P_k = E[V|S_k]$. Thus, if J trades actually take place, then $P_1, \ldots, P_J$ are the prices at which trades occur, and $P_{J+1}$ is some value intermediate between the bid and ask prices at the end of the trading period (this convention is not far from what CRSP does in calculating daily returns). The specification of $H_t^+$ from the preceding paragraph implies that $P_k$ is measurable with respect to $H_k$ which allows us to prove the following proposition.

Proposition 2

The sequence of transaction prices $\{P_k\}$ forms a martingale relative to the specialist's information, $\{S_k\}$ and the public information, $\{H_k\}$.

Proof

From the above, $P_k = E[V|S_k]$. Thus,

$$E[P_{k+1}|S_k] = E[E[V|S_{k+1}]|S_k] = E[V|S_k] = P_k.$$

Since $H_k$ is contained in $S_k$, and since $P_k$ is measurable with respect to $H_k$, the sequence of transaction prices $\{P_k\}$ forms a martingale relative to $\{H_k\}$ as well. Q.E.D.

The above proposition predicts that if, for example, one were to look at daily closing prices, these prices would form a martingale (i.e., price differences would be uncorrelated). The usual explanation for this is that investors, using public information, will drive away any return to this information, and at equilibrium, prices must form a martingale (or a discounted martingale). The use of public information also leads us to the conclusion that prices form a martingale, but here it is competition among risk-neutral specialists, rather than the activity of rational investors that leads to the result.[4]

The result above is actually stronger than the usual statement of the semi-strong form of the efficient markets hypothesis--prices form a martingale relative to all public information and the information known to the specialist. The assumed competition among equally informed specialists implies that there are no profit opportunities arising from the information known by the specialist. Furthermore, at the instant that a trade occurs and the price is announced, the specialist and all outsiders agree on the expected value of V.

The fact that prices form a martingale suggests that if we allow the volume of customers arriving in the trading period to rise without limit, the spread will get small or trading will cease. Since transaction prices form a positive martingale, they converge. This means that after a large number of trades there cannot be large jumps in the price. What we show below is that if there is sufficient trading activity involving both buyers and sellers, then the spread will converge to zero.

<u>Proposition 3</u>

Assume that for all t, the (endogenous) probabilities that a customer arriving at time t buys or sells are bounded away from zero, i.e.,

$P\{Z_t > A_t | S_t\} \geq \alpha > 0$, and $P\{Z_t < B_t | S_t\} \geq \beta > 0$. The sequence of transaction prices $\{P_k\}$ converges. Moreover, $T_k$ (the time of the kth trade) is finite for every k, and the root mean square spread over the first k trades tends to zero as $1/\sqrt{k}$; $\sqrt{(1/k)E[\Sigma(A_j - B_j)^2]} \leq \sqrt{Var(V)/\alpha\beta k}$ .

<u>Proof</u>

From Proposition 1, transaction prices form a martingale. Since these transaction prices are conditional expectations of a positive random variable with finite variance, this martingale converges almost surely. Since the probability of no trade at any arrival is bounded away from one, an infinite number of arrivals will lead to an infinite number of trades almost surely; i.e., $T_k$ is finite almost surely for all k.

Since $P_k$ is a conditional expectation of V, the variance of V exceeds the variance of $P_k$. Taking $P_0$ to be E[V], we have:

$$\mathrm{Var}(V) \geq \mathrm{Var}(P_k) = \mathrm{Var}\left( \sum (P_j - P_{j-1})\right) = E[(\sum(P_j - P_{j-1}))^2]$$
$$= E[\sum(P_j - P_{j-1})^2] + 2E[\sum_{i<j}\sum(P_j - P_{j-1})(P_i - P_{i-1})].$$

The increments in a martingale are uncorrelated and have mean zero, so the second expectation is zero. Hence, for all k:

$$\mathrm{Var}(V) \geq \sum^{k} E[E[(P_j - P_{j-1})^2 | A_j , B_j ]].$$ Some algebraic manipulation shows that

$$E[(P_j - P_{j-1})^2 | A_j , B_j ] \geq \alpha\beta(A_j - B_j)^2$$

Thus, $\mathrm{Var}(V) \geq \alpha\beta\sum^{k} E[(A_j - B_j)^2]$. Q.E.D.

When the condition that the probabilities of buying and selling are both bounded below by a positive number fails to hold, the spread may not shrink to zero. One possibility is that the market may close down entirely, as would occur, for example, if each trader had $\rho$ equal to 1 and the Z's of the informed followed any continuous distribution (c.f. Milgrom and Stokey (1982)). Another possibility is that all trades occur at the bid price, as would happen if it were known to the specialist that each trader had $\rho$ less than one. The bounded probabilities condition can be proved to hold whenever $P\{\rho > 1\}$ and $P\{\rho < 1\}$ are both positive and there are no informed traders, or more generally when the liquidity motive for trade (measured by the dispersion of $\rho$ above and below one) is sufficiently large given the insider trading motive (as determined by the proportion of insiders and the nature of their information). Henceforth we assume that $\rho$ has a continuous distribution with median one so that buying and selling motives are balanced.

The result of Proposition 3, that the spread goes to zero, means that prices "reflect" information. Since the only source of the spread in this model is informational differences, the fact that the spread goes away with many trades suggests that informational differences must go away as well--i.e., prices must eventually reflect essentially all the information of the informed traders. This intuition is sharpened in Proposition 4, where it is shown that the probability that the specialist and an arrival will disagree by much on the expected value of V goes to zero as the number of trades gets large.

Proposition 4.

If the assumptions of Proposition 3 hold, then the expectations of the specialist and the traders converge, i.e., $E[V|S_k] - E[V|F_k]$ converges to zero in probability (where $F_k$ is the information of the trader who makes the kth trade).

Proof

We use the notation and results of Proposition 1 and two general facts from probability theory cited earlier. The suppressed time subscript is now $t = T_k$.

$$A = E[V|C] = E[E[V|F,\rho]|C] = E[E[V|F]|E[V|F] > A/\rho]$$

$$\geqslant E[E[V|F]|E[V|F] > E[V]/\rho].$$

Define D by $D = E[V|F] - E[V]$. Then,

$$\begin{aligned} A - E[V] &\geqslant E[D|D > E[V](1-\rho)/\rho] \\ &\geqslant P\{\rho \leqslant 1\}\ E[D|D > 0] \\ &\geqslant P\{\rho \leqslant 1\}\ P\{D \geqslant \varepsilon\}\varepsilon \\ &= P\{D \geqslant \varepsilon\}\varepsilon/2. \end{aligned}$$

By Propositions 1 and 3, $A_t - E_t[V]$ converges almost surely to zero, so $P\{D_t \geqslant \varepsilon\}$ must also converge to zero for all positive $\varepsilon$. A similar argument using bid prices shows that $P\{D \leqslant -\varepsilon\}$ goes to zero. Thus, $E_k[V|F_k] - E_k[V]$

converges in probability to zero. Also, $E_k[V|S_k] = E_k[V]$ (both are equal to $A_k$ if the customer buys and $B_k$ if he sells). Q.E.D.

Thus, in the long run, the market converges to an informational steady state in which traders all have the same expectation of V and the gains to private information evaporate. The market is also stable in the sense that in the limit, the specialist's inventory of stock and cash is a driftless random walk.

Corrollary 1

The specialist's inventory of stocks tends to a driftless random walk; i.e.,

$$\lim P\{Z_k < B_k \ |H_k\} - P\{Z_k > A_k \ |H_k\} = 0.$$

Proof

Using the results of the Proposition 4, let $P^*$ be the limit of $E[V|S_k]$ and let $\hat{Z} = \rho P^*$. Then,

$$\lim(P\{Z_k < B_k|H_k\} - P\{Z_k > A_k|H_k\}) = P\{\hat{Z} < P^*|S^*\} - P\{\hat{Z} > P^*|S^*\} ,$$

$$= P\{\rho < 1\} - P\{\rho > 1\} = 0.$$ where $S^*$ is the specialist's limiting information. Q.E.D.

That the specialist's inventory will not drift on average is only true in the limit. After some finite number of trades, there is no guarantee that for the next finite number of trades the specialist does not expect a change in his inventory. Only when the market is at the informational steady state is it in equilibrium in the more conventional sense that expected supply equals expected demand.

To summarize, we have shown that a risk neutral competitive (zero profit) specialist will in general set a positive spread between bid and ask price when there are insiders present. Despite the existence of a spread, the fact that the specialist acts competitively implies that the transaction price sequence forms a martingale relative to his information, and hence relative to all public information. If the volume of trade rises without limit, and both buyers and sellers are active, then the spread converges to zero. More specifically, the average spread is bounded on the order of the square root of the number of trades. In the limit, insiders, outsiders and the specialist agree on the conditional expected value of V.

We now turn to an analysis of various environmental assumptions. In particular, we will see that the proportion of insiders and the quality of inside information are directly related to the magnitude of the spread. The dispersion of uninformed liquidity parameters is inversely related to the spread and the thickness of the market (frequency of arrivals of orders) is inversely related to the spread.

To distinguish sharply between insiders and liquidity traders, we assume that the insiders have no liquidity reasons for trading (their $\rho$ is equal to one). We also assume that the specialist has access only to public information; i.e., $S_t = H_t$. If there is an arrival at time t, we can express the reservation price $Z_t$ by:

$Z_t = (1 - U_t)E[V|H_t \vee I_t] + U_t\rho_t E[V|H_t]$, where (as above) $U_t$ is one if an arrival at t is uninformed and zero otherwise (for the succeeding discussion, time subscripts have been dropped to simplify the notation). We now prove some comparative equilibrium results.

<u>Proposition 5</u>

For any given time t, the ask price $A_t$ increases and the bid price $B_t$ decreases when, other things being equal,

(i) the insider's information at time t becomes better (i.e., finer),

(ii) the ratio of informed to uninformed arrival rates at t is increased, or

(iii) the dispersion of the uninformed preference parameter $\rho_t$ (around its median) at t is decreased.

Proof

Define M by $M = E[V|I]$ and let G be the distribution function of the liquidity parameter. Recalling that A and B, the ask and bid, are the smallest a and largest b satisfying

$a \geq E[V|Z > a]$, $b \leq E[V|Z < b]$, then, for ask price A and bid price B,

$AP\{Z > A\} - E[VI_{\{Z > A\}}] \geq 0$, and

$E[VI_{\{Z < B\}}] - BP\{Z < B\} \geq 0$. The left sides can be expanded to

$$E[(1 - U)(A - M)I_{\{M > A\}} + U(A - E[V])(1 - G(A/E[V]))] = E[\phi(A,U,M,G)]$$

$$E[(1 - U)(M - B)I_{\{M < B\}} + U(E[V] - B)G(B/E[V])] = E[\psi(B,U,M,G)].$$

The functions $\phi(.)$ and $\psi(.)$ are concave in M and increasing in U. Also, $\phi(.)$ is decreasing in $G(A/E[V])$ and $\psi(.)$ is increasing in $G(B/E[V])$.

For (i), let A' and B' be the ask and bid prices associated with insider information I' and define M' by $M' = E[V|I']$ where I' is finer than I. For (ii), let A" and B" be the ask and bid prices when the arrival of uninformed is governed by U", where $U''(\omega) \leq U(\omega)$. For (iii), let A"' and B"' be the ask and bid prices when the distribution function of $\rho$ is G"' where

$G'''(x) \leq G(x)$ for $x \leq 1$ and $1 - G'''(x) \leq 1 - G(x)$ for $x \geq 1$.

If I' is finer than I, then $M = E[M'|I]$. This and Jensen's inequality allow us to conclude:

$E[\phi(A',U,M,G)] \geq E[\phi(A',U,M',G)]$. Since $\phi$ is increasing in U, and $U'' \leq U$, $E[\phi(A'',U,M,G)] \geq E[\phi(A'',U'',M,G)]$. Also, since $\phi$ is decreasing in G, $E[\phi(A''',U,M,G)] \geq E[\phi(A''',U,M,G'''))]$. (Similar inequalities hold for the function $\psi(.)$). By the definitions of A', A", and A"' the right hand sides are all nonnegative. But, $A = \inf\{a | E[\phi(a,U,M,G)] \geq 0\}$, so A', A", and A"' all exceed A. The same argument will prove the corresponding bid inequalities. Q.E.D.

Intuitively, the adverse selection problem is worse the greater the fraction of informed traders and the better their information. The specialist is forced to set a higher spread if there are more informed or if they have better information in order to avoid losses. On the other hand, the greater the desire of the uninformed to trade (measured by the dispersion of $\rho$), the easier is the specialist able to make back his losses to informed traders. The zero-profit condition then results in a smaller spread.

It is, of course, possible that in the case of increasing spreads, that the increase will drive out too many liquidity traders and so fail to eliminate the adverse selection. This is identical to the famous "lemons problem" and, as in Akerlof's model (1970), the market can break down entirely with $A_t$ and $B_t$ set to preclude trade.

Proposition 5 does not state that on average, spreads will be smaller for stocks with a small percentage of insiders than for stocks with a larger percentage of insiders. For any given information, spreads are larger, the more informed there are.

There appear to be opposing forces at work. First, as the proportion of insiders increases, the insiders have more of an impact on prices and hence their information is revealed in prices more quickly. Initial spreads are wide, but as the information is revealed, the spread may go to zero more quickly. On the other hand, the more insiders there are, the larger is the initial spread and hence the fewer trades take place. This will tend to make the spread larger for a longer period of time. The same caveat applies to the case when the information that the insiders have is better. The proposition does not claim that average spreads will always be wider when insiders have higher quality information. Once again, there are competing forces. There will be probabilistically fewer trades when insiders have better information, but each trade will be more informative.

Proposition 5 suggests that we might observe the following. For some period of time, we expect firms with a relatively high proportion of insider stockholders to have relatively wide spreads. Just prior to predictable informational events, assuming that insiders have early access to the information, we might expect to see larger spreads than after such an event when the insider advantage is small. Thus, for example, in the month or so before a company's earnings report is made public, proposition 4 would suggest that we would observe larger spreads than after the report is made public. Furthermore, the convergence result might suggest that the spread on average will be smaller immediately prior to the announcement than say one month before the announcement.

While the above propositions do not address the question of average spreads, the following discussion relates average spread to average volume. Suppose one were to watch the bid ask spread for t units of clock time. Let N be the stochastic process describing transactions, i.e., one will see $N_t$ transactions during this time period, and $P_{N_t}$ will be the last transaction price observed. Since this price is a conditional expectation of V, its variance is less than or equal to the variance of V. Recall the expression from Proposition 3: $Var(V) \geq Var(P_{N_t}) \geq E[\sum^{N_t} \alpha\beta\ (A_{T_k} - B_{T_k})^2]$. If the underlying arrival process is Poisson with rate r, then for large r, $N_t$ is on the order of $(\alpha + \beta)rt$. where $\alpha$ and $\beta$ are as in Propostion 3. That is, the average spread is on the order of $1/\sqrt{tr(\alpha + \beta)}$ i.e., on the order of the square root of the average volume per t units of time.

This result reflects the fact that transactions reveal information, albeit imperfectly. Thus, the higher the volume, the more information is conveyed (i.e. becomes public) per unit time and hence the smaller the informational difference between insiders and outsiders. This argument is somewhat different from the usual justification for the observed relationship between volume and spreads. The more common argument is that markets with higher volume are more liquid hence the specialist (who provides liquidity) will receive a lower compensation for the liquidity he provides; i.e. will set smaller spreads. In our analysis, the higher the "liquidity" of the market the more quickly is insider information "revealed," i.e., reflected in prices. This has the effect of lowering the average spread.

## III. A Model with Discounting

The model discussed in the previous section is based on a particular normalization of reservation prices that was mathematically convenient. This normalization took the form of the specialist having a $\rho$ of one, while the median of the $\rho$'s of the traders was one. Another normalization that is of economic interest is the following: the reservation price of an individual arriving at time t is $Z_t^*$ given by $Z_t^* = \exp(-r_t(T_o - t))Z_t$ where $Z_t$ is as defined in the previous section, and $T_o$ is the time of the informational event. The parameter $r_t$ may arise from other unmodeled market opportunities and depends only on time, not on any personal characteristics. The zero profit condition for the specialist now becomes a zero excess return condition and may be stated as (if solutions exist):

$$A_t^* = \exp(-r_t(T_o - t))E[V | S_t^*, Z_t^* > A_t^*]$$

$$B_t^* = \exp(-r_t(T_o - t))E[V|S_t^*,\ Z_t^* < B_t^*].$$

Since the market now being described is merely a renormalization of the one described in section II, it is straightforward to show that $A_t^*$ and $B_t^*$ are given by $A_t^* = \exp(-r_t(T_o-t))A_t$ ; $B_t^* = \exp(-r_t(T_o-t))B_t$ where $A_t$ and $B_t$ satisfy (as above)

$A_t = E[V|S_t,\ Z_t > A_t]$ $\quad B_t = E[V|S_t,\ Z_t < B_t]$

To insure that outsiders have an incentive to be involved in the market, the following hypothesis is offered. Let $\tau$ be a holding period. The expected gross holding period return of someone buying at time t and holding for

$\tau$ periods of time is $\dfrac{E_t[B^*_{t+\tau}]}{A^*_t}$. It is assumed that at any time t,

$$\frac{E_t[B^*_{t+\tau}]}{A_t^*} = e^{i\tau}$$ where i is an exogenously given rate of return.

Although this is implicitly a hypothesis about the exogenous variables, it is stated in terms of market parameters and appears to be testable. The variable i might be taken to be a required return consistent with the risk of the stock. The important limitation such a condition imposes on the data is that i be unrelated to the magnitude of the spread and constant through time. In effect, this assumption defines $r_t$. Since $B^*_{t+\tau}$ is a function of $r_{t+\tau}$ and $A_t^*$ is a function of $r_t$, a terminal condition and the above expected holding period return condition will define $r_t$. The proof of the following proposition is tedious, and is relegated to an appendix.

Proposition 6

Let the expected realizable return of an uninformed trader over the normal holding period be i, i.e., $E_t[\frac{B^*_{t+\tau}}{A^*_t}] = e^{i\tau}$ for all t. Assume that after the informational event at $T_o$, V becomes known so that for

$t \varepsilon [T_o-\tau, T_o]$, $\quad B^*_{t+\tau} = Ve^{-i(T_o-t-\tau)}$

(i.e. $r_{t+\tau} = i$ for $t \varepsilon [T_o-\tau, T_o]$). Then $r_t$, the discount rate at t, is the normal return i plus a premium:

$r_t = i + (n+1)/(T_o-t)\log(k_t)$ where

$$\frac{1}{k_t} = \{E_t[\frac{B_{t+\tau}}{A_t} \frac{B_{t+2\tau}}{A_{t+\tau}} \cdots\cdots \frac{B_{t+(n+1)\tau}}{A_{t+n\tau}}]\}^{1/(n+1)} \leqslant 1$$

and $t+n\tau \varepsilon [T_o-\tau, T_o)$.

The discount rate applied at time t, $r_t$, has a particularly interesting interpretation. Notice that $(\frac{1}{k_t})$ is the expected geometric mean gross return per $\tau$ units of time earned by an investor that follows a strategy of buying and selling every $\tau$ periods of time in a market with no discounting. The log of this is thus the continuously compounded expected return from such a strategy. Obviously, such a return is negative. Recall from the definition of $A_t$ and $B_t$ that the specialists sets the bid and ask so that on average what he loses to the informed is made up by what he gains from the uninformed liquidity traders. Thus, $(n+1)/(T_o-t)\log(k_t)$ ( a positive number) is, in return (per unit time) terms, what the uninformed on average lose to the informed. Thus, $r_t$ represents the expected holding period return, i, plus the return that the uninformed anticipate losing to the informed. Note that $r_t$ depends upon the holding period $\tau$. In particular, n+1 in Proposition 6 is approximately $(T_o-t)/\tau$, and hence $r_t$ is approximately $i + (1/\tau)\log(k_t)$.

The above proposition, with i specified exogenously, closes the model in the sense that $Z_t^*$, $A_t^*$, $B_t^*$ are now specified. The resulting price process will be $\{P_t^*\}$ with $P_t^* = e^{-r_t(T_o-t)} P_t$ where $P_t$ is as specified in the previous section. The observed holding period return will be $\frac{P_{t+\tau}^*}{P_t^*}$. If $T_o-t$ is large relative to $\tau$, then $r_t$ and $r_{t+\tau}$ will be approximately equal, in which case $\frac{P_{t+\tau}^*}{P_t^*}$ will be on average approximately equal to $e^{r_t\tau}$. The observed returns will be larger than i, the hypothesized holding period return, since

$$P_{t+\tau}^* \geq B_{t+\tau}^* \text{ and } P_t^* \leq A_t^*, \text{ and hence } \frac{P_{t+\tau}^*}{P_t^*} \geq \frac{B_{t+\tau}^*}{A_{t+\tau}^*}, \quad \text{which is equal to}$$

$e^{i\tau}$ in expectation. That is, returns calculated by observing transaction prices will always be at least as large as the returns that one could realize by buying at time t and selling at time t+τ.

On the other hand, it is easy to see that the existence of a bid ask spread is less important the longer is the investment horizon. Intuitively, this spread can be amortized over a larger number of periods. To see this, the expected value of the return that can be realized long term is:

$$E_t\left[\left(\frac{V}{A_t^*}\right)\right] = E_t\left[\left(\frac{P_t^*}{A_t^*}\;\frac{P_{t+\tau}^*}{P_t^*}\cdots\frac{P_{t+n\tau}^*}{P_{t+(n-1)\tau}^*}\;\frac{V}{P_{t+n\tau}^*}\right)\right]$$
$$= E_t\left[\left(\frac{P_t^*}{A_t^*}\right)\left(\frac{P_{t+\tau}^*}{P_t^*}\cdots\frac{P_{t+n\tau}^*}{P_{t+(n-1)\tau}^*}\;\frac{V}{P_{t+n\tau}^*}\right)\right]$$

$$= E_t[(\frac{P_t}{A_t}) \; (\frac{P^*_{t+\tau}}{P^*_t} \cdots \frac{P^*_{t+n\tau}}{P^*_{t+(n-1)\tau}} \; \frac{V}{P^*_{t+n\tau}})]$$

Since $P_t \leq A_t$, the above expected return is less than the observed return. If $T_o-t$ is large, however, then $(P_t/A_t)^{1/T_o-t}$ will be close to one, and the long term per period mean return will be close to the observed (from the transaction price sequence) per period return.

These observations may provide some insight into such "anomalies" as the "small firm effect" and the "ignored firm effect." In both cases it may be reasonable to conjecture that informational differences between market participants may be significant. In the case of the small firm effect, it may be the case that insiders hold a larger proportion of the stock. As the results in section II show, this will indicate (other things equal) a larger spread and hence a larger divergence between $r_t$ and i. In the latter case, the lack of public reporting on a firm may imply that there is a larger informational difference between insiders and outsiders. This will also mean a larger spread and hence a greater difference between $r_t$ and i. The above results suggest that the measured "excess returns" are not realizable in a short run basis. Rather, the spread, which represents the expected loss of the uninformed to the informed, leaves an outsider with a "normal" rate of return. In the long run, returns will indeed be larger on average, but these higher returns can only be realized by buying and holding.

## IV. Conclusion

To summarize the above results, we have shown that the existence of traders better informed than the specialist will generate a spread between bid and ask. The resulting transaction price process is a martingale (or submartingale with discounting of the future) relative to all public information. Insiders, however, profit from their information but if we allow trade to continue indefinitely, the profits of late arriving informed traders tends to zero. The model predicts that the average spread should be inversely proportional to the square root of the average volume, and the spread increases with the proportion of insiders and the quality of their information. The existence of a spread implies that returns calculated by examining the transaction price process will overstate the realizable returns. It is shown that if a normal expected return goes to the outsiders, then the return that is measured will be approximately this normal return plus what outsiders expect to lose to insiders.

The analysis shows that prices are informative and hence (loosely speaking) spreads should decline through time. This along with the results of section III should indicate declining excess returns as the time of the informational event is approached. Furthermore, as the spread decreases, the volatility of the stock price should decrease. This entire analysis is of course predicated on their being an informational event, and the existence of individuals with prior access to the information trading based on that information. The model would suggest that further studies of excess returns should control for numbers of insiders, the timing of informational events and, to the extent possible, the quality of inside information relative to publicly available information. It also suggests that the spread itself is of practical importance and should be analyzed in terms of the above variables.

There are several compromises and omissions of the model presented here that should be addressed by further research. First, the restriction of trade to unit lots is unrealistic. Specialists implicitly have a price schedule for various volumes of trade, and a model that allows such schedules could deal with the question of the informational content of volume. It would appear to have the added benefit of predicting a relationship between volume and price volatility. Second, further development of investor behavior facing a specialist market is required. That is, an analysis of other investment opportunities is necessary to develop fully the "required rate of return" notion introduced in section III. Third, the informational source of the spread should be integrated with the other sources of spread due to the fact that the specialist is risk averse (as discussed in Ho and Stoll) to come up with a more complete model of securities pricing. Fourth, while the strategies of the specialist and market participants are dynmamic, the exogenous uncertainty is unchanging. A further development that is in line with other models of the specialist might be to specify some exogenous stochastic process representing the "intrinsic value" of the firm. Finally, an analysis of the welfare effects of the spread and a comparison with those of other possible trading institutions could prove to be interesting.

## APPENDIX

### Proof of Proposition 4

First consider $t \in [T_o - \tau, T_o)$. Then

$$E_t\left[\frac{B^*_{t+\tau}}{A^*_t}\right] = \exp(-i(T_o - t - \tau)) \frac{E_t[B_{t+\tau}]}{e^{-r_t(T_o - t)} A_t} = e^{i\tau}. \text{ That is,}$$

$$r_t = i + \log\left[\left(\frac{1}{\frac{E_t[B_{t+\tau}]}{A_t}}\right)^{T_o - t}\right].$$

For $t+\tau < T_o$, and $t+n\tau \in [T_o - \tau, T_o)$, suppose

$$r_{t+\tau} = i + (n/(T_o - t - \tau))\log(k_{t+\tau}).$$

Then, $$e^{i\tau} = \frac{E_t[\exp(-(i+(n/(T_o-t-\tau))\log(k_{t+\tau}))(T_o-t-\tau))B_{t+\tau}]}{A_t \exp(-r_t(T_o - t))}$$

and $$r_t = i + \frac{1}{T_o - t} \log\left(\frac{A_t}{E_t\left[\frac{B_{t+T}}{k^n_{t+\tau}}\right]}\right)$$

$$= i + \log\left[\left(\frac{1}{E_t\left[\frac{B_{t+\tau}}{A_t \quad K^n_{t+\tau}}\right]}\right)^{1/T_o - t}\right].$$

Now, $$E_t\left[\frac{B_{t+T}}{A_t}\left(\frac{1}{k_{t+\tau}}\right)^n\right] = E_t\left[\frac{B_{t+\tau}}{A_t} E_{t+\tau}\left[\frac{B_{t+2\tau}}{A_{t+\tau}} \cdots \frac{B_{t+(n+1)\tau}}{A_{t+n\tau}}\right]\right]$$

$$= E_t[\frac{B_{t+\tau}}{A_t} \frac{B_{t+2\tau}}{A_{t+\tau}} \cdots \frac{B_{t+(n+1)\tau}}{A_{t+n\tau}}]$$

$$= (1/k_t)^{n+1} \leq 1,$$

since $E_{t+(k+1)\tau}[\frac{B_{t+(k+1)\tau}}{A_{t+k\tau}}] \leq E_{t+k\tau}[\frac{M_{t+(k+1)\tau}}{A_{t+k\tau}}] = \frac{M_{t+k\tau}}{A_{t+k\tau}} \leq 1.$

Thus, $r_t = i + ((n+1)/(T_o-t))\log(k_t)$. The (backwards) induction argument shows that $r_t$ is as claimed. Q.E.D.

## ENDNOTES

1. Copeland and Galai argue that the specialist effectively supplies a call to buyers and a put to sellers. On average, the specialist loses the value of the call and put to informed traders and makes the spread from uninformed traders

2. This model may not be providing the whole story, since as Keim (1981) has noted, much of the excess January return occurs in the first few days of January--before insiders might be presumed to have accurate information.

3. It is convenient, but not necessary to think of the arrivals as forming a Poisson process. Most of the conditions imposed on the arrival process in the propositions will be satisfied if arrivals are Poisson.

4. It should be noted that for small differencing intervals, the martingale property has not held up empirically. As others have argued (1970), this may be the result of bid and ask limit orders piling up on the specialist's book. The result is that transaction prices often appear negatively correlated.

REFERENCES

1. Akerlof, G. A., 1970, The Market for 'Lemons', Qualitative Uncertainty and the Market Mechanism, Quarterly Journal of Economics, 84, 488-500.

2. Arbel, Avner and Paul Strebel, 1981, Neglected Firm Effect and the Inadequacy of the Capital Asset Pricing Model, State University of New York at Binghamton Working paper #81-08.

3. Amihud, Yakov and Haim Mendelson, 1980, Dealership Market, Journal of Financial Economics, 8 31-53.

4. Bagehot, Walter (pseud.), 1971, The Only Game in Town, Financial Analysts Journal, 12-14, 22.

5. Banz, Rolf, 1981, On the Relationship Between Return and Market Value of Common Stocks, Journal of Financial Economics, 9 3-18.

6. Copeland, Thomas and Dan Galai, 1980, Information Effects on the Economics of the Bid Ask Spread, UCLA Working Paper.

7. Fama, Eugene F., 1970, Efficient Capital Markets; A Review of Theory and Empirical Work, Journal of Finance 25 383-417.

8. Garman, Mark B., 1976, Market Microstructure, Journal of Financial Ecnomics 3 257-275.

9. Grossman, Sanford, 1976, On the Efficiency of Competitive Stock Markets When Traders Have Diverse Information, Journal of Finance, 31 573-585.

10. Ho, Thomas and Hans R. Stoll, 1981, Optimal Dealer Pricing under Transactions and Return Uncertainty, Journal of Financial Economics, 9 47-73.

11. Keim, Donald B., 1981, Size Related Anomolies and Stock Return Seasonality: Further Empirical Evidence, University of Chicago Working Paper.

12. Milgrom, Paul R., and Nancy Stokey, 1982, Information, Trade and Common Knowledge, Journal of Economic Theory, 26, 17-27.