Forthcoming *Journal of Finance*

On Dealer Markets Under Competition

Thomas Ho and Hans R. Stoll*

Working Paper 1-80

Graduate School of Business Administration
New York University

and

The Wharton School
University of Pennsylvania

January 1980

Presented at AFA Meetings
December 28, 1979
Atlanta, Georgia

*We thank Marshall E. Blume, Robert Schwartz and David Whitcomb for their helpful comments.

On Dealer Markets Under Competition

by

Thomas Ho and Hans R. Stoll

Securities markets arise to economize on the costs of trading by investors. These costs include the cost of communicating bids and offers, the cost of waiting, and the cost of transferring title. In a call auction market, trading is at periodic time intervals. In a continuous market, trading is possible at any time. A continuous market requires the existence of at least one unexecuted limit order on the buy side and on the sell side. In U.S. markets, these limit orders are often placed by a dealer. Indeed in the over-the-counter (OTC) market, dealer bid and ask quotes constitute the market. On the New York Stock Exchange (NYSE), the best bid or ask price may be that of the specialist or one of the orders in his limit order book. Whether the bid or ask price is set by the dealer or an investor, the factors determining that price are the same. However, because of his proximity to the market the dealer would generally be expected to be able to offer better prices than an ordinary investor faced with higher communication, waiting and transfer costs. Thus, this paper is phrased in terms of dealer, but the broader applicability to any set of investors placing limit orders is implicit.

Much of the theoretical work on dealers [Demsetz (1968), Tinic (1972), Garman (1976), Stoll (1978), Amihud and Mendelson (1979), Copeland and Galai (1979), Ho and Stoll (1979)] has recognized that dealers may face competition from other dealers or investors placing limit orders, but nonetheless has analyzed only a single (representative) dealer. This approach is quite reasonable for the NYSE specialist who has a quasi monopoly position. But it is less applicable when considering other markets such as the OTC market where there are several dealers with equal access to the market. In this paper, we consider the determination of the market bid and ask price when there is more than one dealer in a stock. Our approach is to apply our model [Ho and Stoll (1979)] of an individual dealer operating under return and transactions uncertainty to the case of more than one dealer and the problem of determining the market bid-ask spread.

Garbade and Silber (1979) examine a market in terms of the factors determining the frequency of clearing. They show that dealers increase the frequency of market clearing and reduce the difference between the transaction price and the "true" underlying price of the stock. The size of the effect depends on the willingness of dealers to trade as a function of the difference between the "true" price and the transaction price. Our earlier work [Stoll (1978) and Ho and Stoll (1979)] can be viewed as specifying this function primarily in terms of the inventory risks to the dealer. While their paper offers many useful insights, it is not concerned with the transaction-by-transaction dynamics of a multi-dealer market, something with which we shall deal.

Cohen, Maier, Schwartz and Whitcomb [CMSW] (1978, 1979) deal with the dynamics of determining the market bid and ask price when individuals placing limit orders constitute the market. They ask the question of why the market bid-ask spread does not go to zero when there are many individuals placing limit orders, and they suggest a "gravitational pull" effect that causes incoming orders to trade with existing limit orders rather than to be listed as limit order at better limit prices. However, their model does not fully specify the determinants of the market spread, something we are able to do.

Framework

Due to space limitations, we can only present some results without fully developing the underlying model. That model is one in which each dealer follows a multiperiod strategy that maximizes his expected utility of terminal wealth taking into account not only his own possible future actions but also those of his competitor. The assumptions about the return dynamics for the stock being traded and the process by which transactions in the stock are brought to the market are similar to those in Ho and Stoll (1979). What is different is the existence of at least two competing dealers in a single stock.

Each dealer solves a dynamic programming problem which is considerably more complex than in the single dealer case because in each period each dealer's action depends not only on his own characteristics and inventory position but also on the inventory position of his competitor and his competitor's other characteristics. (The nature of this interdependence is discussed in this paper for the one period case.) The solution of the dynamic programming problem gives an optimal reservaton selling fee, a; and an optimal reservation buying fee, b. The reservation fee is the minimum fee such that the dealer's expected utility of terminal wealth would not be lowered were he to trade at that fee. If p is the true price of the stock in the opinion of the dealer, the dealer would "earn" the fee by buying at $p - b = p_b$, the bid price, and selling at $p + a = p_a$, the ask price. The fees may be negative. We scale $p = 1$ and think of the fees as percentage fees.

As a simplifying assumption, the more formal part of the analysis is restricted to two dealers, A and B, making a market in the same stock. The dealers have homogeneous expectations about the true value of the stock and have the same opportunity to borrow and lend at the risk-free rate. Variables for dealer B are distinguished from variables for dealer A by a superscript "o". Second, in this presentation we restrict ourselves to the one period case which is sufficient to illustrate some important results. The formal solution of the multiperiod problem is left to another paper.

In the one period case the reservation fees can be shown to be

$$\text{Dealer A:} \quad a = R\,\sigma_I^2\,(\tfrac{1}{2}Q-I) \qquad b = R\,\sigma_I^2\,(\tfrac{1}{2}Q+I) \qquad (1.a,b)$$

$$\text{Dealer B:} \quad a^o = R^o\,\sigma_I^2\,(\tfrac{1}{2}Q-I^o) \qquad b^o = R^o\,\sigma_I^2\,(\tfrac{1}{2}Q+I^o) \qquad (1.c,d)$$

where R = coefficient of absolute risk aversion. σ_I^2 = per period variance of the stock's return. Q = fixed dollar transaction size. I = dealer's inventory holding of stock. Basically, the reservation fee represents the cost to the dealer of entering into a transaction that makes his overall portfolio nonoptimal and/or moves him to a less desirable level of risk. The greater the variance of the stock, the greater the dealer's risk aversion, and the larger the transaction size, the larger the fee for given inventory.[1] The fee may be negative if the inventory position makes certain transactions desirable. For the one period case the reservation fees are the same as in Stoll (1978) and similar to the Ho and Stoll (1979) monopoly dealer solution net of monopoly profits. Only in the one period case is each dealer's reservation fee independent of the competing dealer's characteristics.

However, even in the one period case the expected utility of a transaction for each dealer is not independent of the inventory position and other characteristics of the competing dealer. Which dealer makes the next transaction and what amount over the reservation fee can be charged depends on the relative position of the two dealers. We assume that one transaction can occur in the next instance. That transaction is assumed to be a dealer purchase with probability λ; a dealer sale with probability λ; or there may be no transaction with probability $1-2\lambda$. We assume that dealers follow a bidding process for the next transaction; that is, investors are assumed to interrogate dealers to elicit the maximum buying price (p-b) and minimum selling price (p+a) that dealers are willing to bid. This bidding procedure eliminates the usual monopoly profits that would accrue to a single dealer. However, a dealer may earn a profit over his reservation fee -- a producer surplus -- because the dealer is, for example, in an advantageous inventory position with respect to his competitor and can therefore slightly outbid him and still earn a profit. Any dealer who makes a trade at his reservation fee of course realizes no net benefit since the risk he takes on for the one remaining period of return uncertainty is just covered by the fee.

Trading Patterns and Pricing Behavior With Two Competing Dealers

In this section we assume there is no inter-dealer trading. A number of important implications follow immediately from the general framework we have set up.

1. If $I = I^o$ and if a transaction occurs, A will trade if $R < R^o$. Otherwise B will trade. Given identical inventory positions, the less risk averse dealer -- i.e., the "better" dealer -- can offer the lower buying or selling fee.

2. Because inventory positions may differ, the "better" dealer does not always quote the "better" buying _and_ selling fee. That is, a dealer does not have a natural monopoly because he is less risk averse. If A's inventory is sufficiently large, he is less anxious to buy than B and more

anxious to sell. Even if A is the "better" dealer, he is competitive only on the sell side in this case. It is the case that the "better" dealer will capture a larger fraction of the volume of trading in the stock over several time intervals since he is more frequently able to quote the better price.

3. The dealer with the lower reservation fee has no incentive in fact to quote that fee. He will instead quote his competitor's reservation fee (less a small amount). As a result the market bid-ask spread is, in fact, not the inside quote of the reservation prices, but the outside quote of the reservation prices. The market bid-ask spread, s, is

$$s = \max(a, a^o) + \max(b, b^o) \tag{2}$$

The market reservation spread is the minimum spread that could exist without either dealer being worse off. This is

$$s' = \min(a, a^o) + \min(b, b^o) \tag{3}$$

The two are equal only in the special case in which $a = a^o$ and $b = b^o$, which would occur if $R = R^o$ and $I = I^o$. With different inventory positions, one dealer is typically in a better position to buy and the other dealer in a better position to sell. Each dealer periodically earns a producer surplus as a result of his unique ability to absorb a particular risk.

Figure 1 illustrates these points for the case in which $R = R^o$ and dealers differ only because of their inventory positions. Assume A's inventory is $I = 100$ and B's inventory is $I^o = -100$. In this case A is in a better position to sell and can earn a profit by charging B's selling fee. Similarly, B is in a better position to buy and can earn a profit by charging A's buying fee.

While the reservation spread, (3), can be negative, the market spread, (2), is always positive. The lower bound on the market spread in the two dealer case is the reservation spread of the "worst" dealer. The worst dealer is the one with the larger value of R.

Inter-dealer trading, "Gravitational Pull" and the Market Spread

Inter-dealer trading arises if a dealer would rather trade at the existing market bid or ask price than remain with his own bid or ask and take a chance on trading with an incoming market order. Inter-dealer trading is identical to the "gravitational pull" effect that CMSW (1978, 1979) discuss in the context of whether investors place market orders or limit orders. In terms of our model, each dealer must calculate the utility of trading with the other dealer at the posted price compared with the utility of trading with the next market order with probability λ.

For the purpose of illustrating our model consider the one period case in which the dealers are identical except that A has the larger inventory. Suppose $I - I^o \geq 2Q$. This implies the ordering of reservation quotes in Figure 1 and that A is concerned only with the sell side of the market (B is in the better position to buy). Dealer A has two options. Option 1 is to sell to a market order with probability λ to earn a fee of a^o. Option 2 is to sell to dealer B paying to B a fee of Π. For the case of two dealers, we show that if $\Pi = b$, the market buying fee, there will be no inter-dealer trading. However, there is a negotiated fee at which inter-dealer trading would take place.

With one period remaining, A's expected utility under option 1 can be shown to be

$$EU = U(W) + U'\Delta W + \tfrac{1}{2}U''\sigma^2(W) + R(I-I^o)\lambda\sigma_I^2 U'Q \qquad (4)$$

where the first three terms on the RHS represent the expected utility of total end-of-period wealth in the absence of any transactions and given the underlying return dynamics that make uncertain the future wealth. U(W) is the dealer's elementary utility function which has first and second derivatives of U' and U", and W is the dealer's total wealth. The last term of (4) represents the expected utility of the profit earned in a sale transaction. The profit depends on the degree of divergence between the inventory of A and B. If $I = I^o$, the last term of (4) is zero and a sale by A to a market order would not add to A's utility because he is compelled to trade at his own reservation price.

The expected utility of option 2 is derived by substituting the new values of the portfolio components after the inter-dealer transaction into (4), i.e., inventory of A and B become $I - Q$ and $I^o + Q$ respectively. The cash received by A is only $Q(1-\Pi)$. A will trade with B if option 2 yields a greater value for (4). This condition is met if

$$\Pi < \frac{Z}{W} \sigma_I^2 [I - Q(\tfrac{1}{2} + \lambda)] \qquad (5)$$

If inter-dealer trading is allowed only at market quotes, Π is the market buying fee which is set by A, i.e., $\Pi = b$, where b is given by (1.b). Substitution into (5) shows that the condition (5) can then never be met. The reason is quite simple. Even if λ, the probability of selling to a market order, is zero; A will not sell to B and <u>pay</u> the market buying fee to reduce inventory that he is willing to add to at the existing market buying fee. If $\lambda > 0$, A has the additional possibility of selling to a market order and <u>earning</u> a fee. However, there is a <u>negotiated</u> fee at which A would sell to B rather than take a chance on a market order. If that fee is to be positive, it must be the case that $I > Q(\tfrac{1}{2} + \lambda)$.

More Than Two Dealers

If our assumption of two dealers is relaxed, inter-dealer trading is possible even at market prices. The key condition is that the market buying fee is not set by the dealer who is considering selling at the market price. This can be the case with more dealers. Suppose there are three dealers with identical R; and suppose dealers B and C also have identical inventory of $I^o < I$, A's inventory. Therefore, B and C set the market bid and ask price. Because they have the smaller inventory, B and C are in the better position to buy. Because they compete, they will be forced to offer to buy at their reservation bid price. A has the larger inventory and is in the better position to sell. Since there is no competition on the sell side, A is able to quote the higher reservation ask price of B and C rather than his own reservation ask price. Condition (5) can now be met at the market bid since the market bid is not A's reservation bid. Indeed, in this example Π is given by (1d) where the superscript "o" now refers to the dealer setting the market bid price (B or C). Thus, (5) becomes

$$R^o \sigma_I^2(\tfrac{1}{2}Q + I^o) < R \sigma_I^2 [I - Q(\tfrac{1}{2} - \lambda)]$$

and because R = R in this example, this becomes

$$I^o + Q < I - \lambda Q \tag{6}$$

If the "market" inventory (i.e., of B or C) after an inter-dealer transaction, $Q + I^o$, is less than A's expected inventory without an inter-dealer transaction, $I - \lambda Q$, an inter-dealer trade will occur.

The two dealer framework remains applicable when there are more dealers because only two dealer types -- the marginal and infra-marginal -- need be considered on each side of the market. Define the marginal buyer as the dealer with the highest reservation buying price. The marginal seller is the dealer with the lowest reservation selling price.

1. When R is the same for all dealers, the marginal buyer is the one with the smallest inventory and the marginal seller is the one with the largest inventory.

2. Inter-dealer trading by a marginal dealer is possible only when the marginal dealer faces infra-marginal prices. The condition for a gravitational pull effect for the marginal seller when dealers have identical R is (6). The condition for the marginal buyer is the same except that Q is replaced by -Q in (6). As shown earlier, this condition cannot be met in the case of two dealers where the marginal dealer is forced to trade at his own reservation price.

3. Any infra-marginal dealer (who is neither the marginal buyer nor the marginal seller) has zero probability of trading with the next market order, i.e., $\lambda = 0$ for the infra-marginal dealer. This implies a more frequent "gravitational pull" for infra-marginal dealers as reflected in condition (6) in the case of identical R. Indeed, (6) implies inventories of infra-marginal dealers cannot differ by Q or more; for if the difference were Q or more, they would trade with a marginal dealer. Furthermore, under the assumption of identical R, (6) also implies that the inventory of a marginal dealer cannot diverge by 2Q or more from that of any infra-marginal dealer.

4. As in the case of two dealers, the market spread is positive when there are many dealers. Any single dealer has a positive reservation spread. By result three infra-marginal dealers have the same inventories when all R are the same. As a result, they all have the same (positive) reservation spread. If there is only one marginal buyer and one marginal seller, this

reservation spread (of the infra-marginal dealers) will be the quoted market spread. The marginal dealer will earn a profit. In the special case of more than one marginal buyer and one marginal seller (which requires identical I and R for two marginal dealers on each side of the market), the market spread may be less than any reservation spread. This occurs when the inventory of the marginal buyers is less than the inventory of the marginal sellers. None of the marginal dealers can earn a profit. However, the market spread is bounded away from zero because a sufficiently small spread (the value of which we can specify) implies too large a divergence of inventories that by (6) would lead to inter-dealer trading and an increase in the spread. The maximum divergence in inventories is Q by condition (6).

5. The general expression for the market spread is (2) where a is the selling fee of the marginal seller, a^o is the next best selling fee, b is the buying fee of the marginal buyer, and b^o is the next best buying fee. When R differs across dealers, the market spread remains positive for reasons analogous to those already discussed. As before, the market spread is minimized when there are two dealers with the same a and two with the same b and when inventories differ by the maximum amount not leading to inter-dealer trading.

6. While specifying the conditions for inter-dealer trading we have not specified the likely frequency of inter-dealer trading. Under homogeneous expectations among dealers and a single transaction size, one can show that inter-dealer trading will not occur. The reason is simply that dealer inventories can never diverge by enough to satisfy condition (6) (or its analogue under differing R). If a new dealer enters with a large inventory, inter-dealer trading would occur. Similarly, if one allows a dealer to

negotiate transactions of larger size without clearing the quotes of other dealers (as appears to be the case in real dealer markets), he will enter the next period with an unbalanced inventory that would lead to inter-dealer trading.

7. While the size of the market spread depends critically on the number of dealers in the stock, there is nothing to guarantee that there will be sufficient dealers to force the marginal dealer to quote his reservation price. A dealer will make a market for a stock only if the profits he expects to make outweigh the costs associated with setting up a dealer operation. The more dealers in a stock, the less frequent the occasions in which any particular dealer is in position to earn a profit.

Implications of Extending the Time Horizon

Extending the time horizon beyond the one period considered in this paper complicates the solution for the reservation buying and selling fee, even when the problem is restricted to two dealers. In determining his pricing strategy, each dealer must take account of its effect not only on his own future inventory but also the future inventory of his competitor because future profits depend on his position relative to his competitor's position.

It is possible to show that the spread will tend to be wider in this case because each dealer knows that profits arising from a favorable inventory position must erode over time as the flow of transactions causes inventories to be equalized. Thus, each dealer raises his reservation fees (above the one period level) to compensate for the transitory nature of these profits.

Summary and Conclusions

In this paper, we have outlined the problem of determining the reservation bid and ask prices of dealers operating in a competitive multiperiod environment and subject to return uncertainty and transactions uncertainty. The key difference between the competitive case and the one dealer case treated in Ho and Stoll (1979) is that each dealer's pricing strategy depends not only on his own current and expected inventory position and his other characteristics but also on the current and expected inventory and other characteristics of the competitor.

Second, we have examined the determination of the market bid-ask spread in the one period case. Under the bidding equilibrium assumed, each dealer will not generally quote his reservation price -- the price at which he is indifferent between trading or not trading. The dealer with the best reservation price will quote the price of his nearest competitor. When there are two dealers, this implies the market spread is the difference between the highest reservation asking price and the lowest reservation buying price -- i.e., the outside of the reservation quotes. With more than two dealers, the typical spread is reduced relative to the two dealer case because the dealer able to quote the best price has more competitors, thereby limiting his ability to quote a poor price.

Third, we have examined the conditions for inter-dealer trading, i.e., the conditions under which a dealer would prefer to trade at a market bid or ask price rather than post his own quote and wait for an incoming market order. In a multidealer environment, there is an incipient "gravitational pull effect" that limits the divergence of inventories and thereby limits the divergence of reservation prices of dealers. The "gravitational pull effect" is incipient because the flow of incoming orders will be traded in a way that keeps inventories of different dealers from diverging very far. There may be an actual gravitational pull effect for new entrants with an unbalanced inventory. The new entrants will choose to trade with a dealer rather than place their own limit order under conditions specified in the paper.

FOOTNOTES

[1]The variance is the appropriate measure of risk in the expressions for the bid or ask fee because the dealer is not diversified with respect to the transaction in the stock. The covariance between the return on the stock and the rest of the dealer's inventory, of course, shows up in the general expression for the dealer's expected utility of terminal wealth.

REFERENCES

Amihud, Y. and Mendelson, H., 1979, "Dealership Market: Market Making with Inventory," working paper no. 608/79, (May).

Cohen, K., Maier, S., Schwartz, R., and Whitcomb, D., 1978, "Limit Orders, Market Structure, and the Returns Generation Process," Journal of Finance, 33 (June).

Cohen, K., Maier, S., Schwartz, R., and Whitcomb, D., 1979, "Transaction Costs, Order Placement Strategy, and Existence of the Bid-Ask Spread," working paper.

Demsetz, H., 1968, "The Cost of Transacting," Quarterly Journal of Economics, 82 (February).

Galai, Dan and Copeland, Thomas, 1979, "The Economics of the Bid-Ask Spread," UCLA, working paper, September.

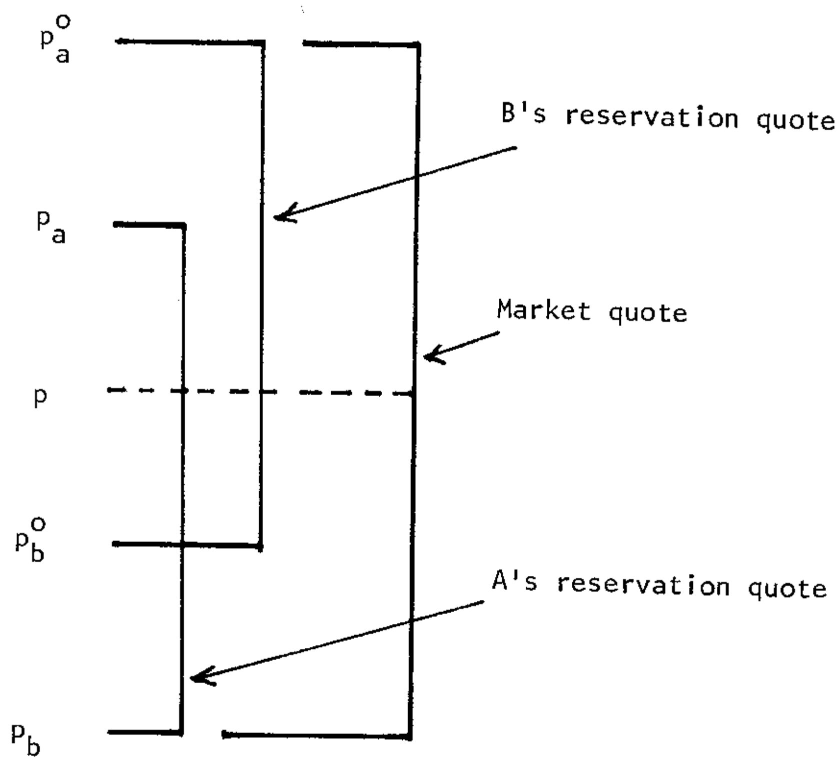
Garbade, K. and Silber, W., 1979, "Structural Organization of Secondary Markets: Clearing Frequency, Dealer Activity and Liquidity Risk," Journal of Finance, 34 (June).

Garman, Mark, 1976, "Market Microstructure," Journal of Financial Economics, 3 (June).

Ho, Thomas and Stoll, Hans, 1979, "Optimal Dealer Pricing Under Transactions and Return Uncertainty," working paper, October.

Stoll, Hans R., 1978, "The Supply of Dealer Services in Securities Markets," Journal of Finance, 33 (September).

Tinic, Seha, 1972, "The Economics of Liquidity Services," Quarterly Journal of Economics, 86 (February).



$p_a = p + a$ $\quad p_b = p - b$ $\quad p_a^o = p + a^o$ $\quad p_b^o = p - b^o$

where a, b, a^o, b^o are given by (1)

A's profit: $a^o - a = p_a^o - p_a$

B's profit: $b - b^o = p_b^o - p_b$

Figure 1