Market Making, the Tick Size, and Payment-for-Order Flow: Theory and Evidence
Author(s): Tarun Chordia and Avanidhar Subrahmanyam
Source: *The Journal of Business*, Vol. 68, No. 4 (Oct., 1995), pp. 543-575
Published by: The University of Chicago Press
Stable URL: http://www.jstor.org/stable/2353145

Tarun Chordia
Vanderbilt University

Avanidhar Subrahmanyam
University of California, Los Angeles

Market Making, the Tick Size, and Payment-for-Order Flow: Theory and Evidence*

We analyze the effects of a finite tick size and the practice of "payment-for-order flow" on market maker competition. Even if the New York Stock Exchange (NYSE) reservation price is superior to its non-NYSE counterpart, brokers may, because of payment-for-order flow, prefer to execute orders off the NYSE floor. In accordance with the implications of the model, empirical analysis suggests that non-NYSE market makers trade a larger fraction of the smaller order sizes and offer fewer price improvement opportunities and that large companies appear to have enhanced price improvement opportunities on the NYSE.

Recently, much concern has been voiced that trading volume on the New York Stock Exchange (NYSE) has been adversely affected by the increasing proportion of trades transacted on the non-NYSE regional exchanges and the "off-floor" market (e.g., the well-known firm of Madoff Securities, Inc.). Further, the merits of the practice of "payment-for-order flow" (where off-floor market makers and non-NYSE exchanges offer cash rebates to brokers in return for order flow) have also been hotly debated.[1] In this article we show that the nature of the pricing and the competition between market makers depends on regulations governing the magnitude of the minimum price change on the NYSE. In the United States, price changes in most financial securities have a lower bound, called the "tick

* We are indebted to Doug Diamond (the editor) and an anonymous referee for comments that have substantially improved this article. We also thank Michael Brennan, Jim Cochrane, Nick Crew, Mike Fishman, Larry Glosten, Joel Hasbrouck, Charles Lee, Ananth Madhavan, L. Shivakumar, Hans Stoll, and seminar participants at the January 1994 meetings of the American Finance Association, Boston, for useful comments and discussions. All remaining errors are our own.

1. See, e.g., *Wall Street Journal* (1994), p. A7. Also see *Wall Street Journal* (1992), p. A1, which indicates that only about 59% of the U.S. stock market transactions in 1991 took place on the NYSE floor, as compared to 76% in 1981.

(*Journal of Business*, 1995, vol. 68, no. 4)

0021-9398/95/6804-0004$01.50

size.'' The tick size is $1/8$ of a dollar for stocks priced at or above \$1, $1/16$ for stocks under \$1 and at or above \$0.25, and $1/32$ for stocks under \$0.25.[2] The concept of a minimum tick size in U.S. financial markets has recently received considerable attention from a regulatory perspective. The issue of whether exchanges should move to a regime with ''decimal'' stock trading has also been taken up to some extent in the popular press.[3]

The tick size has been defended, for example, by Grossman and Miller (1988), who argue that it serves to maintain a minimum level of profits for market makers and, thus, guarantees the provision of liquidity by these agents. In accordance with this intuition, we find that the minimum tick size helps guarantee a minimum level of profits to market makers on the trading floor.[4] This contrasts with the zero-profit assumption made by Glosten and Milgrom (1985), Kyle (1985), and Easley and O'Hara (1987). As documented by Lee (1993), our model also shows that orders submitted to the NYSE may sometimes execute at better than quoted prices. This is because nonspecialist market makers are, at times, able to undercut specialists because of their superior reservation prices.[5] The undercutting is constrained to be in finite amounts because of the presence of the tick size, so that order submitters perceive a randomness about the execution price. Our analysis suggests that the randomness in the execution price would disappear if the floor trader and limit order quotes were disseminated. The institutional structure is such that only the specialist quotes are disseminated.[6]

We next consider competing non-NYSE market makers who can pay brokers for order flow. We show that, even if (i) brokers follow their fiduciary obligation to find the best quoted price for customers, and (ii) NYSE specialists have a lower reservation price, orders may migrate to the non-NYSE market. To understand this result, consider the ask side and a situation where the reservation prices are such that the prices posted by non-NYSE and NYSE market makers are identi-

2. These features are governed by New York Stock Exchange Rule 62 and American Stock Exchange Rule 127.

3. See *Wall Street Journal* (1991), p. C1. This was one of the principal regulatory issues discussed in the Security and Exchange Commission Markets 2000 study.

4. It is assumed in the analysis that on-floor market makers cannot break up orders and execute them at different tick prices to provide traders with any arbitrary average price for the entire order. New York Stock Exchange specialists generally cannot follow this practice because of fixed costs per trade and constraints imposed by the limit order book. See the *New York Stock Exchange Constitution and Rules* (1992), p. 2707.

5. While we explicitly model floor traders, the results would be substantively similar if we were to consider limit order traders as well.

6. Even if the limit order quotes are superior to the specialist quotes, they will remain hidden if the specialist quotes are valid for a different order quantity. McInish and Wood (1992) document that quotes posted by NYSE specialists seldom represent the best limit orders on the book.

cal. In this situation, even if the broker realizes that an NYSE trade may execute at a better than quoted price (because of competition from floor traders), he may still take the order off the exchange floor because the non-NYSE market maker posts the same tick price but pays the broker for order flow. Thus, the practice of payment-for-order flow and the possibility of inferior execution can arise naturally in response to a finite tick size. As the effects described above will not operate in the absence of a finite tick size regulation, our analysis suggests that a move to decimal trading will make the competition for order flow more transparent and orders will flow to the least cost provider of market marking services, thereby positively influencing trading activity.[7]

The model suggests several empirical implications, which we test using Institute for the Study of Security Markets (ISSM) intraday data from 1988 to 1991. We predict that a larger fraction of small trades should be executed on non-NYSE markets and that a larger fraction of small NYSE trades should be filled by the nonspecialist market makers. Our analysis also suggests that the probability of an NYSE trade being filled at better than quoted prices should increase in the number of nonspecialist market makers and that, with payment-for-order flow, there should be fewer price improvement opportunities on non-NYSE markets. These predictions appear to be borne out by the data.

The remainder of this article is structured as follows. The next section provides the basic model, which incorporates the effects of both inventory and asymmetric information. Section II analyzes competition between NYSE and non-NYSE market makers in the presence of a finite tick size, while Section III examines empirical implications obtained from the model. Section IV concludes. A lemma or a proposition that is stated without proof follows directly either from straightforward calculations or from the discussion preceding its statement. The proofs of other lemmas and propositions are provided in Appendix A.

I. The Basic Setting

A. Symmetric Information and Inventory Effects

We first illustrate our basic model by considering the case of symmetric information and a market maker who faces inventory carrying costs, and we then consider the effect of asymmetric information in the next

7. In other related work, Glosten (1991) addresses some issues related to those in this article but does not consider the effect of a finite tick size. Bernhardt and Hughson (1994) consider the effect of discrete prices on dealer competition but do not examine NYSE versus non-NYSE market-maker competition and the issue of payment for order flow.

subsection. Following Madhavan and Smidt (1991) and Zabel (1981), we consider a risk-neutral market maker whose inventory carrying costs increase quadratically in the amount of inventory he or she holds. At time 0, the market maker trades shares of a single risky security that will pay a liquidating dividend $\tilde{v}$ at a future date 1.[8] For simplicity, $\tilde{v}$ is assumed to have an ex ante mean of zero. The market maker posts bid-ask prices that depend on the size of the order, q.

The market maker's cash holding at time 0 is denoted K, and his pretrade inventory level at time 0 is denoted by I. The market maker faces no short-selling constraints or transaction costs in maintaining a market in the designated security. Furthermore, he can freely borrow or lend at the risk free rate, which is assumed to be zero. We assume for now that only liquidity traders are present in the market, that is, there is no information asymmetry and thus, no adverse selection costs. Liquidity traders trade because of their unmodeled liquidity needs. We assume that the liquidity order q is drawn from $[q_{\min}, q_{\max}]$, where $0 < q_{\min} < q_{\max}$.[9]

Let $W_1(I) = I\tilde{v} + K$ denote the market maker's wealth at date 1 when his inventory is I, and further suppose that for a particular inventory level I, the market maker faces an inventory carrying cost of cI^2. Then, for the market maker to achieve his reservation utility, it is necessary that the expected utility from accepting an order is

$$EU[\tilde{W}_1(I)] = K - cI^2. \tag{1}$$

This is the minimum expected utility that the market maker will accept when trading with investors, in the sense that anything less will cause him to stop functioning as a market maker. In other words, (1) gives the market maker's reservation, or "autarky" utility. We will now compute the market maker's reservation bid and ask prices.

Suppose that the market maker sells q shares at the ask price, a. Then

$$\tilde{W}_1(I - q) = (I - q)\tilde{v} + qa + K. \tag{2}$$

Similarly, when the market maker buys q shares at the bid price, b, we have

$$\tilde{W}_1(I + q) = (I + q)\tilde{v} - qb + K. \tag{3}$$

By (2), the market maker's date 1 expected utility from selling q shares is

$$EU[\tilde{W}_1(I - q)] = K + qa - c(I - q)^2. \tag{4}$$

8. One can conceptualize date 1 as the date at which a dividend announcement is made, which removes all uncertainty about earnings.

9. The minimum order size, $q_{\min}$, is typically 100 shares, and the maximum, $q_{\max}$, is typically 10,000 shares. Orders larger than 10,000 shares are executed in the upstairs market.

Similarly, by (3), the market maker's date 1 expected utility from buying q shares is

$$EU[\tilde{W}_1(I + q)] = K - qb - c(I + q)^2. \tag{5}$$

For the market maker to achieve his reservation utility, we equate (4) and (5) to (1) to solve for the ask price, a, and the bid price, b, respectively. The reservation bid and ask prices are given by

$$a = c(q - 2I) \tag{6}$$

and

$$b = -c(q + 2I). \tag{7}$$

The cq term in (6) and (7) reflects the effect of trading q shares on the bid-ask prices. The market maker is ex post rational, and the bid-ask prices reflect the posttrade effects. Any trade of q shares that moves the market maker away from his pretrade inventory position causes the bid-ask prices to be different from the conditional expected values. If a trade results in a posttrade carrying cost equal to the autarky carrying cost (i.e., if $c(I - q)^2 = cI^2$ or $c(I + q)^2 = cI^2$), then by (6) and (7) the bid or ask prices are set equal to the expected liquidation value of zero.

The $2cI$ term in (6) and (7) represents the effects of the existing inventory, I, on the bid-ask prices. The term arises because a larger (positive) inventory leads to a decrease in the reservation price on both the bid and the ask sides. Note also that the ask (bid) price need not be monotonically increasing (decreasing) in the carrying cost coefficient. For example, from (6), if $2I > q$, a larger carrying cost results in a *lower* ask price. This is because, if I is large, a marginal increase in c leads to a greater increase in the cost of carrying the pretrade inventory than that of carrying the posttrade inventory; this leads to a decrease in the ask price. The reverse intuition holds for $2I < q$. An analogous discussion applies on the bid side.

B. *Inventory and Asymmetric Information Effects*

We now introduce traders with private information and, in a limited manner, analyze the effects of asymmetric information on the bid and ask prices. Our approach is similar to that of Glosten and Milgrom (1985), except that we model inventory effects and allow for trades of any size. Biais and Hillion (1994) take an approach similar to ours in a model of the interaction between stock and options markets.

For tractability, we need to impose further structure on the model; thus we now assume that $\tilde{v}$ has a two-point support, specifically, that $\tilde{v} = \{+\gamma, -\gamma\}$ with equal probability, where $\gamma > 0$. Risk-neutral informed traders learn the realization of the random variable $\tilde{v}$ perfectly; we denote this realization by v. Following Biais and Hillion (1994), we assume that the realization of the size (but not the sign) of the liquidity order, q, is observed by the informed trader as well as the market maker.

Let $a^p(q)$ and $b^p(q)$ denote the posted ask and bid prices for the order quantity q, respectively. We assume that the liquidity traders are equally likely to submit buy or sell orders. The market maker is unable to observe the source of the order flow, in the sense that, when he observes a sell or a buy order of q, he cannot discern whether the order was initiated by an informed trader or a liquidity trader. The market maker attempts to infer $\tilde{v}$ given the sign of the trade q. Given the above structure, all trades by the informed trader that do not equal either zero or q will be out-of-equilibrium trades.[10] The order submitted by the informed trader will therefore endogenously equal $+q$ if $v = \gamma$ and $a^p(q) < \gamma$; $-q$ if $v = -\gamma$ and $b^p(q) > -\gamma$, and zero otherwise.

Let $\theta = B$ represent a buy and $\theta = S$ represent a sale. Since the order size provides no information (just as in Biais and Hillion [1994]), $E[\tilde{v}|q, \theta] \equiv E[\tilde{v}|\theta]$. This is the expected liquidation value given the buy or the sell order. Let $\alpha(\theta)$ denote the market maker's assessed probability of the trade q with sign θ, being initiated by an informed trader. Then

$$\left.\begin{aligned} E[\tilde{v}|B] &= \alpha(B)\gamma \\ &\text{and} \\ E[\tilde{v}|S] &= -\alpha(S)\gamma. \end{aligned}\right\} \tag{8}$$

Note that $\alpha(\theta)$ depends on the relative proportion of informed traders in the market, as well as the probability of an informed trader buying shares (as opposed to selling shares). Though $\alpha(\theta)$ is an endogenous variable, it can easily be characterized as follows.

Suppose for the moment that the ask price $a^p(q)$ is less than γ. Then the informed trader with positive information (i.e., who observes $v = +\gamma$) will buy, and $\alpha(B)$ can be defined in terms of exogenous parameters. Let the proportion of informed traders in the population be α. If the ex ante proportion of informed traders with positive information is ½, then, since the liquidity traders are also equally likely to buy or sell, $\alpha(B) = \alpha$ and $\alpha(S) = \alpha$. If the ask price quoted by the market maker is greater than γ, however, then the informed trader with positive information will not always buy, and there may be an equilibrium in which an informed investor randomizes between participating and not participating. An analogous characterization applies for the bid side.

For concreteness, and purely for illustrative purposes, we assume

10. This claim is supported by the out-of-equilibrium market-maker belief that any order of size greater than q is initiated with a probability of one by an informed trader. An order of size less than q is evidently suboptimal for the informed trader.

in the remainder of this section that all market makers are identical, in the sense that they all have the same inventory and carrying cost parameters, and that there are at least two such market makers who bid in a Bertrand-like fashion for the order flow.[11] This ensures that the reservation prices are the equilibrium bid-ask prices as well. The following lemma provides expressions for the reservation bid and ask prices in the presence of inventory and adverse selection costs, while imposing restrictions on exogenous parameters to ensure that the informed trader adopts the pure strategy of always participating.

LEMMA 1. Suppose that all market makers have the same inventory and carrying cost parameter and that there are at least two such market makers. Then, if

$$\gamma > \max\left[\frac{c(q-2I)}{1-\alpha}, \frac{c(q+2I)}{1-\alpha}\right], \tag{9}$$

the bid-ask prices (the reservation prices) for an order quantity q are

$$b^p(q) = -\gamma\alpha - c(q + 2I) \tag{10}$$

and

$$a^p(q) = \gamma\alpha + c(q - 2I). \tag{11}$$

The bid-ask spread is

$$a^p(q) - b^p(q) = 2\gamma\alpha + 2cq. \tag{12}$$

Condition (9) stipulates that the bid-ask reservation prices of lemma 1 are valid for a given inventory, I, and a given order size, q, as long as $a^p(q) < \gamma$ and $b^p(q) > -\gamma$. These constraints ensure that the informed trader always participates in the market. As pointed out earlier, if the above constraints are violated, the informed trader may adopt a randomized strategy across participating and not participating. Analysis of this complicated scenario abstracts from the main issues addressed by this article, so we focus on the parameter space under which the informed trader adopts a pure strategy of participating.

Setting $\alpha = 0$ in (10) and (11) gives the bid and ask prices that should prevail in the presence of inventory effects alone, as given by (6) and (7). Substituting $c = 0$ in (10) and (11) gives us the bid and ask prices due to information effects alone. Note that due to asymmetric information, the ask (bid) price is increasing (decreasing) in the market maker's assessed probability that the incoming order is from an informed trader.

The remainder of this article will use the bid-ask prices derived in lemma 1 to analyze institutional features which govern market making.

11. We consider heterogeneous market makers in the next section.

II. Competition and the Tick Size

In this section, we address two institutional features of market making that have recently received much attention: the prevalence of a minimum tick size in securities trading and the emergence of the practice of "payment-for-order flow." Throughout the analysis, we only consider the ask side; the analysis for the bid side is analogous.

A. *The Tick Size*

We will now consider heterogeneous market makers. Let us first examine the pricing function when the market maker is constrained to set bid-ask prices in units of a tick size, which we denote T. A crucial issue in our analysis is whether a market maker can split an order across two different tick prices to obtain any average price for a particular order and thereby avoid the constraints imposed by the tick size regulation. New York Stock Exchange specialists are, in general, precluded from following this practice because of per-trade fees imposed on them by clearing houses and constraints imposed by the public limit order book.[12] Accordingly, we assume that an order must be executed at a single price.

We consider a specialist and a nonspecialist market maker (a floor trader) who competes with the specialist for order flow. While the analysis incorporates nonspecialist market making by considering a single floor trader,[13] the basic intuition would remain unaltered if we were to model limit order submitters instead. Let the inventory cost parameter and the inventory level of the specialist be c_s and I_s and those of the floor trader be c_f and I_f.[14] The inventory cost parameters are common knowledge. We let $a_s(q; c_s, I_s)$ and $a_f(q; c_f, I_f)$ denote the specialist's and the floor trader's reservation ask prices, respectively.

The sequence of events in the model is as follows:

1. The size and sign of the liquidity order q are realized. The specialist observes only the size of the liquidity order, q. The informed trader observes the realization of $\tilde{v}$ and q perfectly.

12. On the NYSE, 70%–80% of the trades in actively traded stocks do not involve the specialists and are instead the result of public order meeting public order (see, e.g., Madhavan and Smidt 1991). The limit order book, therefore, imposes a nontrivial constraint on the specialist. These details were obtained from personal communication with Gene Savin, an NYSE specialist. An exception occurs in the case of "dividend-capture" trades, in which the trades are structured to account for the exact amount of the dividend per share.

13. Introducing more than one floor trader increases the complexity of the model without adding much intuition.

14. Our model is intended to describe intraday phenomena that span a few minutes of trading time. We therefore do not allow for trading between the specialist and the floor traders to correct temporary inventory imbalances across these agents. The idea is that, in an intraday setting, it is unrealistic to expect the inventory positions of agents to be optimally balanced at all times.

2. The utility maximizing specialist posts his quote for the given order q taking into account the fact that the quotes are constrained by the tick size. The specialist is forced by institutional regulation (e.g., by way of monitoring by floor officials) to allow a floor trader to receive the order if the floor trader quotes a lower price when the order is exposed to the floor.
3. The utility maximizing floor trader, also constrained by the tick size, has an opportunity to bid for the order before it is executed. The floor trader receives the order if and only if his quote is superior to that of the specialist.

Note that the informed trader's strategy is unchanged from the previous section by way of the constraint that, to avoid being identified, he must mimick the liquidity traders whenever he submits non-null trades. That is, the strategy of the informed trader who observes $v = +\gamma$ is to submit $+q$ when the posted ask price is less than γ and not to trade otherwise (the bid side strategy is analogous).

We now characterize the equilibrium strategies of the specialist and the floor trader. The equilibrium of our game has the flavor of a Stackelberg equilibrium because the specialist moves first, taking into account the best response of the floor trader to his own strategy.

SPECIALIST'S STRATEGY. Let N be the smallest integer such that $NT > a_s(q; c_s, I_s)$. Thus, NT is the tick price just above the specialist's reservation ask price. To characterize the specialist's optimal quote-setting strategy, we need to delineate his beliefs. Suppose that, from the specialist's perspective, the floor trader's inventory I_f is distributed uniformly in $[-L, L]$. This implies that the specialist's prior distribution on the floor trader's reservation price a_f^* is uniform in $[\alpha\gamma + c_f q - 2c_f L, \alpha\gamma + c_f q + 2c_f L]$. The cumulative distribution function (c.d.f.) placed by the specialist on the floor trader's reservation ask price, a_f^*, is then given by $G[x] = (x - \alpha\gamma - c_f q + 2c_f L)/4c_f L$.

The specialist will quote an ask price $(N + k)T$ that conditional on his beliefs, maximizes his expected utility. In order to ensure that the informed trader always participates in the market—that is, he submits $+q$ when he observes $v = +\gamma$ and vice versa—we will impose the condition that

$$\alpha\gamma + 2c_f L + c_f q_{\max} \leq \gamma - 2T, \tag{13}$$

that is, that the upper bound of the floor trader's reservation ask price is no greater than $\gamma - 2T$. Given that the reservation ask price increases in q (see [11]), condition (13) ensures that it is never optimal for the specialist to quote an ask price greater than or equal to γ. Given his beliefs the specialist will solve the following problem in choosing his ask price, $(N + k)T$:

$$\max_k EU[W_1(I, k)|B]G[(N + k - 1)T] + EU[W_1(I - q, k)|B]\{1 - G[(N + k - 1)T]\}, \tag{14}$$

where the argument k in W_1 emphasizes the dependence of the expected utility on the quoted ask price. The first term in (14) is the expected utility of the specialist if he does not fill the order multiplied by the probability of not getting the order when he quotes an ask price of $(N + k)T$. The specialist does not receive the order if the floor trader's reservation ask price is $(N + k - 1)T$ or lower, enabling the floor trader to take the order by quoting $(N + k - 1)T$. The second term in (14) is the expected utility of the specialist if he does fill the order multiplied by the probability of getting the order. Thus, the specialist chooses that price $(N + k)T$ which maximizes his expected utility formed on the basis of the distribution of the floor trader's reservation ask price a_f^* (conditional on the sign of the trade). Purely for expositional purposes, we assume that, when the solution to (14) is not unique, the specialist chooses the lowest value of k that maximizes his utility, subject to the rule that the posted price be at least as high as his reservation price for the order.

FLOOR TRADER'S STRATEGY. If $a_f(q; c_f, I_f) \leq (N + k - 1)T$, then given the specialist's posted quote of $(N + k)T$, the floor trader's best response is to quote $(N + k - 1)T$ and receive the order at this price. If $a_f(q; c_f, I_f) > (N + k - 1)T$, the floor trader is unable to, and therefore does not, bid for the order.

The following lemma summarizes the above strategies.

LEMMA 2. The equilibrium strategies of the specialist and the floor trader are

1. *Specialist.*—Quote an ask price of $(N + k^*)T$, where k^* solves the maximization problem in (14). If k^* is not unique, choose the lowest quote $(N + k^*)T$ that is above the reservation price for the order.
2. *Floor Trader.*—If $a_f(q; c_f, I_f) \leq (N + k^* - 1)T$, quote $(N + k^* - 1)T$ and receive the order at this price. If $a_f(q; c_f, I_f) > (N + k^* - 1)T$, do not bid for the order.

Using parameter values described in table 1, we now provide a numerical example in order to intuitively characterize these strategies. Note from (11) that the reservation ask price is increasing in q and that $\partial^2 a/\partial q \partial c > 0$, so that as long as two market makers have the same inventory, the reservation ask price (for a given q) of a market maker who faces a higher inventory cost parameter always lies above that of a market maker who faces a lower inventory cost parameter. An

TABLE 1 Parameters of Example

Variable Description	Variable	Value
Specialist inventory carrying cost	c_s	.0005
Floor trader inventory carrying cost	c_f	.0008
Non-NYSE inventory carrying cost	c_m	.001
NYSE specialist inventory level	I_s	0
Floor trader inventory level	I_f	55
Non-NYSE inventory level	I_m	30
Probability of an informed trade	α	.10
Uniform distribution support	γ	.5
Specialist knows floor trader's inventory	I_f^*	55

empirically realistic scenario is one in which floor traders face higher carrying costs than the specialists on the floors of organized exchanges, that is, $c_f > c_s$.[15] Then, if $I_s = I_f$, the floor trader will be unable to compete with the specialist for order flow (since $a_f(q; c_f, I_f) > a_s(q; c_s, I_s)$). However, the ask price decreases in a market maker's inventory level. Thus, if $I_f > I_s$, it is possible that the floor trader may have a reservation price which dominates that of the specialist for some order quantities, that is, $a_f(q; c_f, I_f) \leq a_s(q; c_s, I_s)$ for some $I_f > I_s$. Figure 1 presents one such reservation price function. In the following discussion based on this figure, we assume (for simplicity) that the specialist knows exactly the inventory level of the floor trader, that is, $I_f^* = I_f$.

In Figure 1, point *X* corresponds to the order quantity at which the two reservation price functions cross. For order quantities smaller than *X*, the floor trader's reservation price is superior, while the opposite is true for order quantities larger than *X*. Note that, except for region *E*, the specialist, based on his equilibrium strategy obtained from solving (14), quotes an ask price that is at a tick just above his reservation price. In region *E*, the specialist optimally quotes a price two ticks above his reservation price and still gets the order in equilibrium because the floor trader's reservation ask price is higher than *NT*.

Optimal order execution requires that orders flow to the least cost provider. However, the presence of the tick size may defeat this objective. In regions *A* and *C* in figure 1, orders flow to the specialist even though the floor trader has lower reservation prices in these regions. Consider region *C*. Orders arriving at the specialist desk will be filled by the specialist at an equilibrium price of ¼, which is the best quote available on the floor. Because of the tick size, both the floor trader and the specialist are forced to offer the same prices in regions *A* and *C*, even though the floor trader has superior reservation prices.

15. This may be the case due to the special privileges bestowed upon the specialist by the NYSE. In any case, our intuition is valid even if $c_f \leq c_s$.

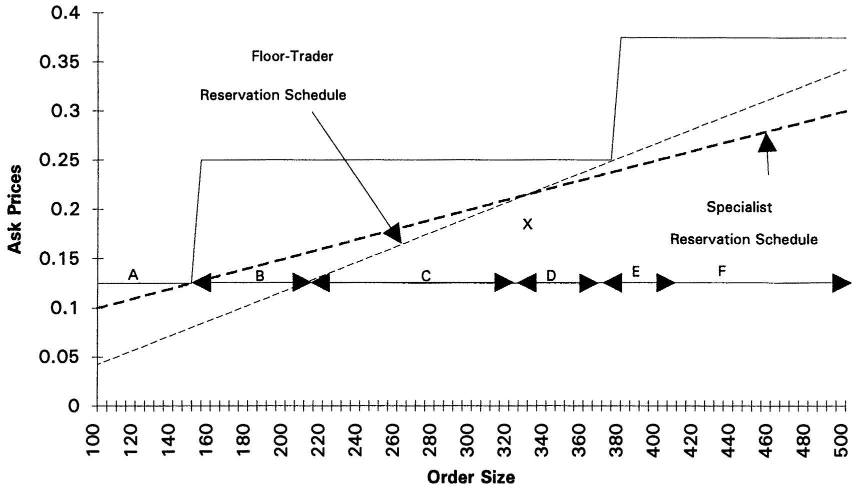

FIG. 1.—The specialist and the floor trader. This figure describes the order flow between the specialist and the floor trader in the presence of the tick size. Optimality in order execution requires that orders flow to the least-cost provider; that is, orders smaller than those corresponding to point X in the figure should flow to the floor trader, and larger orders should flow to the specialist. However, because of the tick size, only orders in region B flow to the floor trader.

In region B of figure 1, the order flows to the floor trader. The price quoted by the specialist is 1/4. However, when the order reaches the floor, it is taken up by the floor trader who is able to quote 1/8 (as per his equilibrium strategy in lemma 2) owing to his superior reservation price. Thus, orders that flow to the exchange floor are sometimes filled at better than quoted prices. From the point of view of order submitters or empiricists who do not observe the inventory levels of the specialists or floor traders, it will appear that an order submitted to the floor is randomly filled at the quoted price or at better than the quoted price. The above discussion can be formalized as the following proposition.

PROPOSITION 1. Depending upon the inventory levels of the specialists and floor traders, orders submitted to the exchange floor are sometimes filled at better than quoted prices. Furthermore, from the perspective of the order submitter, the price at which the order is filled is random.

Formally, suppose that there exist order quantities q^* and q^{**} such that $a_s(q^*; c_s, I_s) = (N + k^* - 1)T$ and $a_f(q^{**}; c_f, I_f) = (N + k^* - 1)T$ and that $a_f(q; c_s, I_s) < a_s(q; c_f, I_f)$ for all $q \in [q^*, q^{**}]$, where k^* is the optimum from (14) for all $q \in [q^*, q^{**}]$. Then all order sizes $q \in [q^*, q^{**}]$ flow to the floor trader and are executed at a price $(N + k^* - 1)T$, which is lower than the posted price $(N + k^*)T$.

Proposition 1 is consistent with Lee (1993), who shows that trades on the NYSE often occur "within the spread." In our model, this phenomenon occurs because of the presence of nonspecialist market makers. It is evident that the probability of obtaining a better than quoted price is (weakly) increasing in the number of such market makers.

Note that it is the institutional structure of the NYSE that lends a randomness to the execution process. The specialist is obligated to quote a price, which must be above his reservation price. He thus leaves himself open to undercutting by the floor traders, once the order reaches the floor. If instead the rules stipulated that the "inside quote" (i.e., the best price taking into account the reservation prices of all agents willing to take the order) should be posted rather than the specialist's quotes, the undercutting would not obtain. Also, it is because of a discrete grid that order submitters perceive a randomness in the execution price (without a discrete grid, floor traders would only have to undercut the specialist by a "penny"—which would again eliminate the randomness).

Since the specialist quotes an ask price which is above his reservation price, he will earn rents in excess of that indicated by his reservation utility. The proposition below formalizes this observation.[16]

PROPOSITION 2. In a finite tick size regime, the NYSE market makers earn rents in excess of that indicated by their reservation utility for almost all order sizes for which they obtain the order.

In contrast to the zero profit assumption made by Glosten and Milgrom (1985), Kyle (1985) and Easley and O'Hara (1987), proposition 2 is consistent with Grossman and Miller (1988) who argue that the tick size serves to maintain a minimum level of profits for market makers. Coffee (1990) states that "pricing securities only in terms of ⅛ of a point intervals essentially amounts to a system of mandatory commissions for dealers and specialists alike" and goes on to suggest that a move to decimal stock trading would make the competition for order flow "more visible" to the ultimate customers. It may be worthwhile to provide a measure of the gains to market makers because of the presence of the tick size regime. Figure 2 presents these gains as a percentage of the dollar value of a transaction, for various transaction sizes. As the figure suggests, these profits can be extremely large in certain cases.

In the following subsection, we consider the effect of the presence of off-floor market makers on optimal order execution.

16. In the proposition, the qualification "almost" is intended to exclude the knife-edge cases where the reservation price corresponding to a quantity exactly equals a multiple of T.

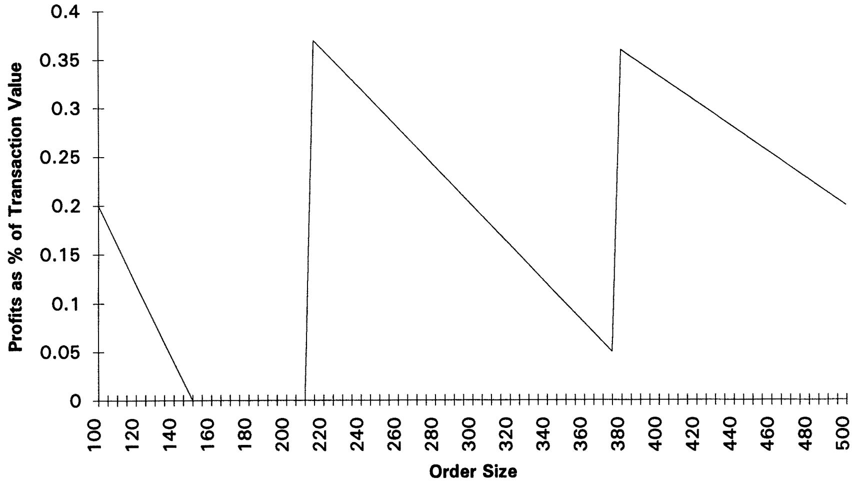

FIG. 2.—Specialist profits because of tick size. This figure shows the amount of the excess rents earned by the specialist because of the presence of the tick size. The rents are in terms of the percentage of the transaction value and are a function of the order size.

B. *Payment-for-Order Flow*

In this section, we assume that a buy order q is executed by a "broker," and consider the implications of introducing off-floor market makers who can "pay for order flow." In accordance with institutional reality, we again assume that specialists post quotes for the realized order size and are obligated to either trade at their quotes or improve on their quotes.[17] We consider a single off-floor (non-NYSE) market maker and let c_m denote the inventory cost parameter, I_m the inventory level, and $a_m(q; c_m, I_m)$ the reservation ask price of this market maker. Note that in figure 3, for orders larger than 120 shares, the reservation price function of the non-NYSE market maker is inferior to those of the specialist and the floor trader. Our analysis to follow suggests that, even if the off-floor market maker is the least efficient provider of market making services (as in the above example), he can cause orders to migrate away from the NYSE floor if he can "pay for order flow."

It has been speculated that the existence of "satellite" markets that practise "payment-for-order flow" may have distorted brokers' incentives to obtain the best price for their customers. However, a popular counterargument is that this is an irrelevant issue since brokers have

17. These features obtain under a quote-driven trading mechanism as described in Madhavan (1992). Equity exchanges and dealer markets in the United States generally conform to the quote-driven system.

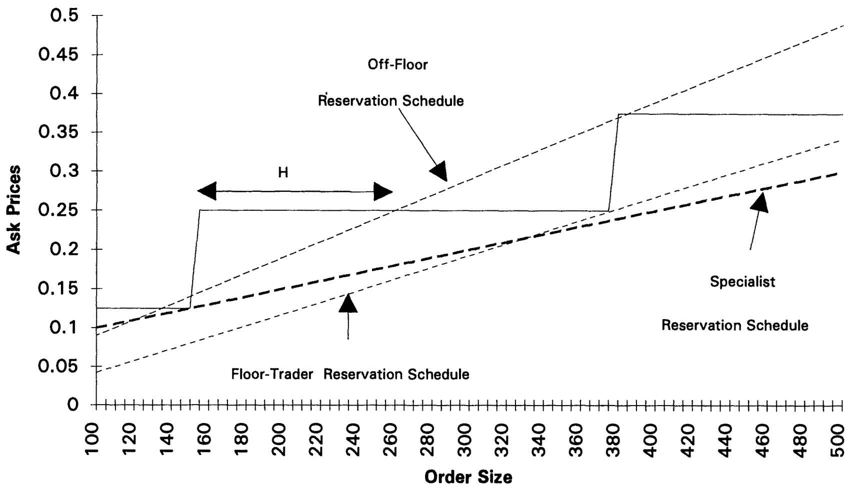

FIG. 3.—The NYSE specialist, floor trader, and the non-NYSE market maker. This figure describes the order flow between the NYSE specialist, the NYSE floor trader and the non-NYSE market maker in the presence of the tick size. Owing to payment-for-order flow, orders in region *H* may flow to the non-NYSE market maker even though he has the least favorable reservation ask prices.

a fiduciary obligation and are required, as per the Securities Act, to obtain the best price for their customers.[18] We argue below that, even if this obligation can be enforced, the institutional structure of the market—in particular, the presence of a finite tick size and the practice of payment-for-order flow—can prevent customers from obtaining the best possible price for their orders.

For an order q^*, let N^* be an integer such that $N^*T > a_m(q^*; c_m, I_m)$. If the order is executed off the NYSE floor at an ask price of N^*T, the total gain to the off-floor market maker before making the payment to the broker is $\Delta \equiv N^*T - a_m(q^*; c_m, I_m)$ (the difference between the upper tick price closest to the non-NYSE reservation price and the reservation price), part of which can be passed on to the broker as payment-for-order flow. If the order is executed on the NYSE floor, the broker does not obtain any payment but faces the possibility of obtaining a better than quoted price from the floor trader (see proposition 1). The broker's objective is to maximize his revenue from order execution. The posted quotes for the order quantities on and off the NYSE floor can be observed by a "regulatory body" (e.g., the Securities and Exchange Commission), and, under the threat of an appropriately large penalty, the broker is constrained to execute at the best *posted* price or an even better price.

18. See, e.g., Peake (1992).

We now characterize the equilibrium strategies of the specialist, the off-floor market maker, the floor trader, and the broker.

SPECIALIST'S STRATEGY (in the Presence of Off-Floor Market Maker). The specialist once again quotes an ask price, $(N + k)T$, to maximize his expected utility. Let $H[\cdot]$ represent the c.d.f. placed by the specialist on the off-floor market maker's reservation ask price, a_m^* (via his inventory I_m). Once again, to ensure that the informed trader always participates, we assume that $\alpha\gamma + 2c_f L + c_f q_{\max} \leq \gamma - 2T$. In the presence of the off-floor market maker, the specialist solves the following maximization problem:

$$\max_k EU[W_1(I, k)|B]G[(N + k - 1)T]H[(N + k)T] \\ + EU[W_1(I - q, k)|B]\{1 - G[(N + k - 1)T]H[(N + k)T]\}. \tag{15}$$

Thus, the specialist again chooses his optimal quote based on his prior distributions of the inventories of the floor trader and the off-floor market maker. The first term in (15) is the expected utility of the specialist when he does not receive the order multiplied by the probability of not getting the order. (The specialist will not receive the order if either the floor trader can post a better quote or if the off-floor market maker can match the specialist quote and offer payments to the broker.) The second term in (15) is the expected utility of the specialist when he does receive the order multiplied by the probability of receiving the order. Again, we assume that when the solution to (15) is not unique, the specialist chooses the lowest value of k that maximizes his utility, subject to the rule that the posted price be at least as high as his reservation price for the order.

OFF-FLOOR MARKET MAKER'S STRATEGY. If the specialist quotes $(N + k)T$ and $a_m(q; c_m, I_m) < (N + k)T$, then the off-floor market maker quotes $(N + k)T$ and makes a onetime commitment to offer a payment of an ϵ arbitrarily close to zero for the order.[19] If $a_m(q; c_m, I_m) \geq (N + \mathrm{k})T$, the off-floor market is unable to, and therefore does not, bid for the order.

FLOOR TRADER'S STRATEGY (in the Presence of Off-Floor Market Maker). Given the quote of $(N + k)T$, if the off-floor market does not capture the order (i.e., if $a_m(q; c_m, I_m) \geq (N + k)T$), and if $a_f(q;$

19. While side payments to brokers in reality are typically 1–2 cents per share, in our 1-period model, it is optimal for the broker to take the order to any market maker that offers side payments, however small the payment. Thus, a side payment of an ϵ arbitrarily close to zero will result in the order flowing off-floor as long as the specialist's quote is not better than that of the off-floor market maker. To preclude bargaining issues, we assume that the off-floor market maker makes a onetime commitment to offer ϵ. If bargaining between the broker and the off-floor market maker were permitted, standard arguments from the bargaining literature can be used to characterize the outcome of the game.

$c_f, I_f) \leq (N + k - 1)T$, then the floor trader quotes and receives the order at a price of $(N + k - 1)T$. Otherwise, the floor trader is unable to, and therefore does not, bid for the order.

BROKER'S STRATEGY. The broker's objective is to maximize revenue from order execution, subject to the constraint that the price at which the order is executed be less than or equal to the posted specialist quote. Since only the off-floor market maker offers a side payment, the broker will direct the order to this market maker so long as the specialist's quoted price is not lower than the price offered off-floor.

The following lemma summarizes the above strategies.

LEMMA 3. The equilibrium strategies of the specialist, the off-floor maker, the floor trader, and the broker, are:

1. *Specialist*.—Quote an ask price of $(N + k^*)T$, where k^* solves the maximization problem in (15). If k^* is not unique, choose the lowest quote $(N + k^*)T$ that is above the reservation price for the order.
2. *Off-floor market maker*.—If $a_m(q; c_m, I_m) < (N + k^*)T$, then quote a price that is the same as that of the specialist and commit to a onetime offer of a side payment of ϵ arbitrarily close to zero. If $a_m(q; c_m, I_m) \geq (N + k^*)T$, do not (you cannot) quote a price for the order.
3. *Floor trader*.—If the off-floor market maker does not receive the order and $a_f(q; c_f, I_f) \leq 9N + k^* - 1)T$, then bid $(N + k^* - 1)T$ and receive the order. If $a_f(q; c_f, I_f) > (N + k^* - 1)T$, do not (you cannot) bid for the order.
4. *Broker*.—If the off-floor market maker quotes the same or lower price than $(N + k^*)T$, send the order to him and receive the payment of ϵ. Else, execute the order on the exchange floor.

Continuing with the numerical example introduced in the previous subsection, consider the region H in figure 3. Both the NYSE specialist and the non-NYSE market maker quote a price (based on their equilibrium strategies in lemma 3) of $1/4$, but the non-NYSE market maker offers a payment to the broker to obtain the order. Now, even if the floor trader may offer a superior price of $1/8$ if the order were to reach the floor, the equilibrium strategy of the broker dictates that he take the order off the NYSE floor to obtain the payment. However, superior execution obtains on the NYSE since there is the possible benefit of a better than quoted price. The broker has discharged his fiduciary responsibility by obtaining the best *quoted* price for his customer. Note that the situation described in figure 3 is more likely to occur for small order sizes as the two reservation price functions diverge for large order sizes under the likely scenario in which $c_s < c_m$ (see [11]). This is consistent with anecdotal evidence that, typically, the orders executed off the NYSE floor tend to be small (see Coffee 1990).

The above discussion can be formalized as a proposition.

PROPOSITION 3. Suppose that the conditions stated in proposition 1 hold. In addition, suppose that there exists a $q\dagger$ such that $q^* < q\dagger < q^{**}$ and $(N + k^*)T = a_m(q\dagger; c_m, I_m)$ and that only the non-NYSE market makers can pay for order flow. Then all orders $q \in (q^*, q\dagger)$ flow off the NYSE floor and are executed at a price $(N + k^*)T$; whereas, in the absence of the non-NYSE market maker, the order would have been executed on the NYSE floor at a price of $(N + k^* - 1)T$.

Thus, from proposition 3 and figure 3 we see that orders may flow off the NYSE floor even though the NYSE reservation prices are superior. This problem can be reduced, though not eliminated, by posting the "inside" quote (the best quote from among the specialists and the nonspecialist market makers) on the NYSE. Thus, disseminating the floor trader quote will hinder the ability of non-NYSE market makers to pay for order flow. However, orders may still flow off the NYSE floor because of presence of the tick size. Recall that in figure 1 some orders were taken up by the NYSE specialist even though the floor trader had superior reservation prices. Since the two effects described above do not obtain in the absence of a tick size, our analysis suggests that a move to a trading regime without a mandated minimum tick size will result in orders flowing to the least cost provider of market making services. This is summarized in the following proposition, which is stated without proof.

PROPOSITION 4. In the finite tick size regime, orders may not flow to the least-cost provider of market making services. Superior execution would obtain in a trading regime that does not mandate a minimum tick size.

Thus, our model predicts that, with decimal stock trading, not only would orders flow to the least cost provider but a larger portion of the total orders would flow to the NYSE floor.[20]

III. Empirical Implications and Tests

We now use the ideas developed in the earlier sections to obtain empirical implications. From (11), if $c_m > c_s$ (an empirically realistic scenario), then the NYSE and non-NYSE reservation price functions will

20. Throughout the analysis in this subsection, we have assumed that traders cannot choose the venue of their order execution, which is consistent with the assumed price-inelastic behavior of liquidity traders. In App. B, we provide an example in which results similar to those in this subsection obtain even when customers can inform the broker where to execute the order.

diverge for large order sizes. Our model thus predicts that the non-NYSE market makers should be in a better position to siphon away the smaller orders. Also, because non-NYSE market makers pay for order flow, they are less likely to execute orders "within the spread." Further, if favorable reservation prices of nonspecialist market makers result in orders being executed within the spread to increase in the number of market makers. Finally, as already discussed, figure 1 and the fact that $\partial^2 a/\partial c \partial q > 0$ in (11) suggest that, when the nonspecialist market maker has the more favorable inventory exposure, his reservation price function should intersect that of the specialist from below. This implies that nonspecialist market makers will be in a better position to take the small orders.

The foregoing discussion thus suggests the following testable implications:

HYPOTHESIS 1. A larger fraction of the small orders will be traded on non-NYSE markets.

HYPOTHESIS 2. With payment-for-order flow, a smaller fraction of the non-NYSE trades, relative to NYSE trades, will execute at better than quoted prices.

HYPOTHESIS 3. The probability of an NYSE trade being executed at better than quoted prices increases in the number of market makers.

HYPOTHESIS 4. A larger fraction of the small orders on the NYSE will be filled by the nonspecialist market makers.

In the following subsections, we use intraday data to test each of the above hypotheses.

A. Data

The ISSM data set used in this study contains intraday data on bid-ask quotes and prices and quantities of each stock traded on the NYSE and the other regional exchanges. Owing to the size of this data set, only 24 stocks were considered. Twelve large and visible companies were chosen to represent a wide range of industries. Twelve other companies with Standard Industrial Classification (SIC) codes as close to the original "large" firms were also chosen. Specifically, each company in the second group was chosen to have an SIC code as close as possible to a corresponding company in the first group (starting with the first digit of the code), subject to the requirement that a firm in the large firm group had to be at least twice as large as the corresponding firm in the small firm group, as measured by the market capitalization at the end of 1988. The market capitalization and SIC codes were obtained from Center for Research in Security Prices (CRSP) tapes. Table 2 provides a list of the companies, their SIC codes, their stock prices and their market values at year-end 1988.

While hypotheses 1 and 2 can be tested directly, hypotheses 3 and 4 require data on the number of nonspecialist market makers for each

TABLE 2 **Description of the Sample Firms**

Kind of Industry	Firm	SIC Codes	Cap	Price
Financial:				
Large firm	Citicorp	6711	8,239	26
Small firm	Chemical	6025	2,962	31
Automobiles:				
Large firm	GM	3711	25,553	84
Small firm	Chrysler	3711	6,011	26
Chemical:				
Large firm	Dow	2812	16,325	88
Small firm	Olin	2810	1,050	51
Computer:				
Large firm	IBM	3573	72,165	122
Small firm	DEC	3573	12,457	98
Drugs:				
Large firm	Merck	2831	22,753	58
	Pfeizer	2834	9,592	58
Small firm	Abbott	2834	10,835	48
	A. H. Robins	2834	622	26
Electric:				
Large firm	GE	3634	40,407	45
	Chevron	2911	15,651	46
Small firm	Emerson	3621	6,845	30
	Ashland	2911	1,970	34
Oil:				
Large firm	Exxon	2911	57,551	44
	Mobil	2911	18,755	46
Small firm	Murphy	2911	1,003	30
	Tesore	2911	166	12
Retail:				
Large firm	Wal-Mart	5311	17,741	31
Small firm	Alexanders	5311	336	68
Telecommunications:				
Large firm	AT&T	4811	30,870	29
Small firm	SNE Telecom	4811	1,694	54

NOTE.—This table provides a list of the companies in the sample, their Standard Identification Codes (SICs), their market capitalizations in millions of dollars (Cap), and their stock prices per share, rounded to the nearest integer (Price), at year-end 1988.

firm. This information, however, is difficult to obtain. Our discussions with NYSE research economists and a specialist indicate that there is a strong positive relationship between the market capitalization of a firm and the number of floor traders. We therefore use firm size as a proxy for the number of nonspecialist market makers in the ensuing empirical analysis.

The ISSM data, which includes all trading days for the years 1988–91, is filtered as follows:

1. Opening batch trades not preceded by an opening quote are excluded since they occur only on the NYSE and do not have a regional counterpart.

2. Trades reported out of sequence, those with special settlement conditions, and those following the daily closure are not considered.
3. Trades on Instinet and the Cincinnati Stock Exchange are excluded since a large fraction of these trades represent "crosses" that are not intermediated by market makers.[21]
4. Following Lee (1993), only NYSE quotes that are eligible for the Best-Bid-or-Offer (BBO) calculation are used as reference quotes.
5. In accordance with the convention employed by Lee and Ready (1991), any quote in the 5 seconds preceding the trade is ignored in favor of a previous quote, except when the trade occurs within 5 seconds of the NYSE opening quote, in which case the opening quote is used.

B. *New York Stock Exchange and non–New York Stock Exchange Market Makers*

Table 3 depicts the total NYSE and non-NYSE trades for different order quantities and different bid-ask spreads. It is evident that non-NYSE market makers captured a substantially larger fraction of the smallest trades (100–400 shares). This fraction decreases in the order quantity. For instance, in 1988–89 non-NYSE market makers accounted for 49.37% of the smallest trades versus 9.35% of the largest trades. The proportions for 1990–91 are even more striking. These results appear to be consistent with hypothesis 1, in the sense that a larger fraction of the small orders tend to be traded on non-NYSE markets.

Even more dramatic is the fact that non-NYSE markets offer fewer price improvement opportunities for the smaller trades. This can be seen in Table 4, which lists the fraction of trades that execute at better than quoted prices. For the smallest trades, NYSE specialists offer significantly greater price improvement opportunities than their counterparts elsewhere. For example, in 1988–89, 68.1% of NYSE trades were executed within the quoted spread, whereas this is true for only 40.4% of non-NYSE trades. For larger trades, non-NYSE market makers often provide more price improvement (this effect is not as noticeable in 1990–91). However, the larger trades constitute a small fraction of the total trades; therefore, NYSE price improvement opportunities appear to be superior overall. Consistent with the notion that inventory holding costs increase with order quantity, we find that price improvement opportunities on the NYSE decline with the size of the trade. Note, though, that price improvement opportunities in non-NYSE

21. The results were substantively unchanged when Cincinnati and Instinet trades were included.

TABLE 3 Total Number of Trades in NYSE and non-NYSE Markets

Spread (in Tick Size)	Trade Size						
	100–400	500–1,000	1,000–1,900	2,000–4,900	5,000–9,900	10,000+	All
A. 1988–89:							
1/8:							
N	526,190	194,787	214,731	189,681	86,459	60,017	1,271,865
NN	513,547	80,490	55,966	22,001	6,899	6,027	684,930
%NN	49.39	29.24	20.67	10.39	7.39	9.13	35.00
1/4:							
N	326,041	136,886	157,303	138,693	60,484	40,706	860,113
NN	324,665	72,137	57,307	24,269	6,316	4,437	489,131
%NN	49.89	34.51	26.70	14.89	9.46	9.83	36.25
3/8:							
N	32,131	15,353	18,801	15,517	6,153	3,752	91,707
NN	24,898	7,228	4,325	1,495	402	360	38,708
%NN	43.66	32.01	18.70	8.79	6.13	8.75	29.68
1/2+:							
N	8,810	4,917	6,359	5,031	1,971	1,195	28,283
NN	7,909	1,714	852	245	83	79	10,882
%NN	47.31	25.85	11.82	4.64	4.04	6.20	27.79
All:							
N	893,172	351,943	397,194	348,922	155,067	105,670	2,251,968
NN	871,019	161,569	118,450	48,010	13,700	10,903	1,223,651
%NN	49.37	31.46	22.97	12.10	8.12	9.35	35.21

B. 1990–91:							
1/8:							
N	900,727	337,025	358,108	279,604	115,514	75,216	2,066,194
NN	1,081,872	159,841	111,666	38,852	9,738	7,492	1,409,461
%NN	54.57	32.17	23.77	12.20	7.77	9.06	40.55
1/4:							
N	472,712	205,656	231,448	189,694	80,081	54,387	1,233,978
NN	597,013	107,805	80,880	28,702	6,821	4,438	825,659
%NN	55.81	34.39	25.90	13.14	7.85	7.54	40.09
3/8:							
N	34,149	17,613	20,493	15,515	5,770	3,557	97,097
NN	35,897	7,399	4,512	1,355	342	256	49,761
%NN	51.25	29.58	18.04	8.03	5.60	6.71	33.88
1/2+:							
N	6,197	3,637	4,369	3,282	1,174	685	19,344
NN	7,431	1,232	771	220	46	33	9,733
%NN	54.53	25.30	15.00	6.28	3.77	4.60	33.47
All:							
N	1,413,785	563,931	614,418	488,095	202,539	133,845	3,416,613
NN	1,722,213	276,277	197,829	69,129	16,947	12,219	2,294,614
%NN	54.92	32.88	24.36	12.41	7.72	8.37	40.18

Note.—This table provides the distribution of trades in NYSE and non-NYSE markets for different bid-ask spreads and different order quantities; N denotes NYSE trades, NN denotes non-NYSE trades, and %NN denotes the percentage of non-NYSE trades.

TABLE 4 Fraction of Trades within the Quoted Spread

Spread (in Tick Size)	Trade Size						
	100–400	500–1,000	1,000–1,900	2,000–4,900	5,000–9,900	10,000+	All
A. 1988–89:							
1/4:							
N	.668*	.621	.593	.573	.560	.522	.617*
NN	.388	.611	.701*	.718*	.563	.440	.477
3/8:							
N	.793*	.689	.655	.615	.591	.533	.693
NN	.580	.774	.773*	.767*	.652	.619*	.646
1/2+:							
N	.741*	.643	.613	.578	.541	.476	.641*
NN	.501	.663	.635	.608	.602	.608	.541
All:							
N	.681*	.629	.600	.577	.562	.521	.625*
NN	.404	.627	.705*	.720*	.569	.456	.490
B. 1990–91:							
1/4:							
N	.665*	.647	.646	.663	.682*	.672*	.660*
NN	.461	.618	.682	.678	.565	.436	.512
3/8:							
N	.808*	.760	.738	.706	.699	.612	.754
NN	.674	.767	.778	.752	.656	.592	.699
1/2+:							
N	.769*	.741	.730	.699	.657	.503	.727*
NN	.641	.716	.717	.715	.596	.434	.657
All:							
N	.676*	.658	.655	.666	.683*	.667*	.667*
NN	.475	.629	.678	.681	.570	.445	.524

NOTE.—This table provides the distribution of within-the-spread trades in NYSE and non-NYSE markets for different bid-ask spreads and different order quantities; N denotes the NYSE and NN denotes the non-NYSE markets.
* Significant at the 5% level based on the paired sign test.

markets first increase and then decrease with order size. This result is consistent with payment-for-order flow by non-NYSE market makers for small order sizes. Also supported is hypothesis 2, in which, with payment-for-order flow, a smaller fraction of orders executed on non-NYSE markets will execute at better than quoted prices.

Market makers will naturally prefer to execute trades at higher bid-ask spreads for a fixed level of adverse selection and inventory costs. Table 3 shows that non-NYSE market makers attract the largest order flow, 36.25% in 1988–89, when the bid-ask spread is $0.25, and 40.55% in 1990–91, when the spread is $0.125. This could be a result of a trade-off between higher profits with higher spreads and the fact that higher spreads may themselves result from increased adverse selection.

Tables 3 and 4 together show that non-NYSE market makers execute a much larger fraction of the smaller orders and offer far fewer price improvement opportunities for these orders. This suggests that the small orders that are not executed on the NYSE are not exposed to competition from the nonspecialist market makers on the NYSE who may often offer superior prices.

C. *New York Stock Exchange Specialists and Nonspecialist Market Makers*

Table 5 shows the distribution of NYSE trades in large and small firms (see table 2 for a description of the sample). The fraction of small firm trades tends to increase in the bid-ask spread, possibly because of greater adverse selection in small firms. Table 6 documents the fraction of NYSE trades that execute at better than quoted prices. Large firm trades typically occur within the spread more often than small firm trades. For instance, in 1988–89, 66.01% of all trades in large firms were within the quoted spread, as opposed to 54.59% of those in small firms. Similar results hold for 1990–91. Assuming that firm size is a reasonable proxy for the number of nonspecialist market makers, the findings documented in tables 5 and 6 support hypothesis 3, in which the probability of an NYSE trade being executed at better than quoted prices increases in the number of market makers. Furthermore, the fraction of trades that occur within the quoted prices decreases monotonically in the order quantity for all bid-ask spreads, providing evidence for hypothesis 4, in which a larger fraction of the smaller orders on the NYSE will be filled by floor traders.

Harris (1993) shows that large firms have high-priced stocks and small firms have low-priced stocks.[22] Since the tick size is more likely to be binding for low-priced stocks, it is reasonable to expect more

22. This is confirmed in table 2, where in 9 out of 12 cases, large firms have higher stock prices than small firms.

TABLE 5 **Total Number of NYSE Trades in Large and Small Firms**

Spread (in Tick Size)	Trade Size 100–400	500–1,000	1,000–1,900	2,000–4,900	5,000–9,900	10,000+	All
A. 1988–89:							
1/8:							
L	417,763	147,965	167,308	152,283	70,893	48,896	1,032,108
S	108,427	46,822	47,423	37,398	15,566	11,121	266,757
%S	20.61	24.04	22.08	19.72	18.00	18.53	20.97
1/4:							
L	230,995	93,805	112,723	102,312	45,488	30,725	616,048
S	95,046	43,081	44,580	36,381	14,996	9,981	244,065
%S	29.15	31.47	28.34	26.23	24.79	24.52	28.38
3/8:							
L	14,916	6,928	9,721	8,560	3,464	2,111	45,700
S	17,215	8,425	9,080	6,957	2,689	1,641	46,007
%S	53.58	54.88	48.30	44.83	43.70	43.74	50.17
1/2+:							
L	4,287	2,895	3,971	3,139	1,189	636	16,017
S	4,523	2,122	2,388	1,892	782	559	12,266
%S	51.34	43.16	37.55	37.61	39.68	46.78	43.37
All:							
L	667,961	251,493	290,723	266,294	121,034	82,368	1,682,873
S	225,211	100,450	103,471	82,628	34,033	23,302	569,095
%S	25.21	28.54	26.05	23.68	21.95	22.05	25.27

B. 1990–91:							
1/8:							
L	759,402	271,056	299,556	237,142	98,072	62,792	1,728,020
S	141,325	65,969	58,552	42,462	17,442	12,424	338,174
%S	15.69	19.57	16.35	15.19	15.10	16.52	16.37
1/4:							
L	370,759	151,212	179,147	148,908	63,288	43,579	956,893
S	101,953	54,444	52,301	40,786	16,793	10,808	277,085
%S	21.57	26.47	22.60	21.50	20.97	19.87	22.45
3/8:							
L	22,654	12,390	15,479	11,769	4,258	2,488	69,038
S	11,495	5,223	5,014	3,746	1,512	1,069	28,059
%S	33.66	29.65	24.47	24.14	26.20	30.05	28.90
1/2+:							
L	4,079	2,670	3,617	2,745	986	512	14,609
S	2,118	967	752	537	188	173	4,735
%S	34.18	26.59	17.21	16.36	16.01	25.26	24.48
All:							
L	1,156,894	437,328	497,799	400,564	166,604	109,371	2,768,560
S	256,891	126,603	116,619	87,531	35,935	24,474	648,053
%S	18.17	22.45	18.98	17.93	17.74	18.29	18.97

Note.—This table provides the distribution of trades in large (more nonspecialist market makers) and small firms (fewer nonspecialist market makers) for different bid-ask spreads and different order quantities (see table 2); L denotes large firm trades, S denotes small firm trades, and %S denotes the percentage of small firm trades.

TABLE 6 Fraction of NYSE Trades within the Quoted Spread

Spread (in Tick Size)	Trade Size: 100–400	500–1,000	1,000–1,900	2,000–4,900	5,000–9,900	10,000+	All
A. 1988–89:							
1/4:							
L	.711*	.673*	.632*	.603*	.592*	.554*	.656*
S	.564	.509	.493	.488	.463	.421	.518
3/8:							
L	.819*	.731*	.694*	.637*	.613	.553	.717*
S	.770	.654	.613	.588	.564	.508	.669
1/2+:							
L	.767	.663	.613	.572	.522	.472	.642
S	.717	.616	.613	.589	.569	.481	.639
All:							
L	.719*	.677*	.636*	.605*	.592*	.553*	.660*
S	.601	.536	.518	.507	.482	.435	.546
B. 1990–91:							
1/4:							
L	.697*	.684*	.666*	.682*	.708*	.692	.687*
S	.557	.549	.576	.589	.585	.591	.567
3/8:							
L	.825*	.779*	.749*	.714*	.708	.619	.766*
S	.798	.720	.707	.680	.675	.596	.737
1/2+:							
L	.837	.772	.748	.700	.652	.502	.753
S	.721	.697	.695	.684	.632	.484	.670
All:							
L	.706*	.692*	.674*	.685*	.707*	.686*	.694*
S	.584	.566	.589	.598	.593	.590	.584

Note.—This table provides the distribution of within-the-spread trades for large and small firms for different bid-ask spreads and different order quantities; L denotes large firms and S denotes small firms.

* Significant at the 5% level based on the paired sign test.

"within the spread" trades in small stocks. However, our analysis in fact suggests the *opposite* (hypothesis 3) and appears to be supported by the data (table 5). Tables 5 and 6 together suggest that the presence of nonspecialist market makers results in a greater probability of orders being transacted at better than quoted prices. Thus, it appears to be the presence of these other market makers that results in more price improvement opportunities on the NYSE as opposed to non-NYSE markets.

IV. Summary and Concluding Remarks

This article examines the effects of a finite tick size and the practice of payment-for-order flow on the issue of competition between NYSE and non-NYSE market makers. In the presence of a finite tick size, order submitters to the NYSE perceive a randomness in the execution price because specialists, who are obliged to post quotes at all times, leave themselves open to undercutting by the nonspecialist market makers and the limit order book. This randomness in the execution price would be eliminated if the inside quote (and not just the specialist quote, as is now the case) were widely disseminated.

Further, suppose that, in conformance with institutional reality, non-NYSE market makers can "pay for order flow." Then our analysis suggests that, in the presence of a finite tick size, brokers have a significant incentive to take the orders off the NYSE to obtain the payment even if they are aware that NYSE orders may sometimes execute at better than quoted prices. In this manner, brokers are able to technically fulfill their fiduciary obligation (of finding the best posted price) to customers while obtaining the payments offered by the off-floor market makers.

Our empirical results support the claim that orders do not flow to the least cost provider of market making services. Due to payment-for-order flow, non-NYSE orders are not exposed to competition from the nonspecialist market makers who often provide superior trade prices. A significant policy issue is whether all orders should be aggregated on the NYSE floor before being allowed to migrate off the NYSE floor, that is, whether competition on a single market offers superior execution than fragmented markets. Peake (1992) argues that forcing all market makers (NYSE and non-NYSE) to compete for order flow at a single location would provide the best execution for investors. Our analysis suggests that, while dissemination of the "inside" NYSE quote will reduce the opportunities for payment-for-order flow, a move to decimal trading (i.e., a trading regime without any mandated minimum tick size) will make the competition for order flow more transparent, and orders would then flow to the least cost provider of market-making services. This should help promote trading volume on the NYSE floor.

As a final point, the existence of the tick size also has implications for the relative rarity of the 1/8th quotes on Nasdaq (Christie and Schultz 1994), which has led to allegations of collusion on the part of dealers on the NASDAQ system. (While higher odd-eighth quotes are also not as prevalent as the even-eighth quotes, the Christie and Schultz [1994] paper suggests that it is the lack of prevalence of the 1/8 quote that raises the strongest suspicion of collusion among Nasdaq market makers.) To illustrate, consider a situation in which the quoted bid and ask prices on a stock are 20 and 20 1/4, respectively. Suppose a market maker wishes to capture order flow by increasing the bid to 20 1/8. Unfortunately, this leads to an increase in his inventory, which he can lay off by reducing his ask price to 20 1/8. But given that his bid-ask spread has been reduced to zero, he has no incentive to offer the odd-eighth quote of 20 1/8. Note that it is the 1/8 tick size restriction that forces the effective spread to zero; the result has no relation to the idea that market makers collude to keep a minimum spread of 1/4.

Appendix A

Proofs

Proof of Lemma 1. The date 1 wealth, when the market maker purchases q shares, is

$$\tilde{W}_1(I + q) = (I + q)\tilde{v} - qb + K, \tag{A1}$$

and, in the event of market maker sale, the date 1 wealth becomes

$$\tilde{W}_1(I - q) = (I - q)\tilde{v} + qa + K. \tag{A2}$$

We now set the expected utility of wealth from (A1) and (A2) equal to $E(U[\tilde{W}_1(I)]|\theta)$:

$$E[(u[\tilde{W}_1(I + q)])|S] = IE(\tilde{v}|S) + K - cI^2, \tag{A3}$$

and

$$E[(U[\tilde{W}_1(I - q)])|B] = IE(\tilde{v}|B) + K - cI^2. \tag{A4}$$

From (A3) and (A4), we have

$$qE[\tilde{v}|S] - qb - c(2qI + q^2) = 0, \tag{A5}$$

and

$$-qE[\tilde{v}|B] + qa + c(2qI - q^2) = 0. \tag{A6}$$

Condition (9) ensures that the informed trader always participates by submitting a buy order when he observes γ and vice versa. Using (8), equations (A5) and (A6) can then be solved for the bid and ask prices in lemma 1. Q.E.D.

Proof of Proposition 1. For an order $q \in [q^*, q^{**}]$, the optimal strategy of the specialist is to quote $(N + k^*)T$, which is the solution to the maximization problem in (14). Note that the utility maximizing specialist will not post a quote that is lower than his reservation ask price $a_s(q; c_s, I_s)$. Given the equilibrium strategy of the floor trader in lemma 2, the order, on reaching the NYSE floor, will be filled by the floor trader at $(N + k^* - 1)T$. Q.E.D.

Proof of Proposition 2. In a regime with finite ticks, the utility-maximizing specialist quotes an ask price that is larger than (or equal to) the reservation price, that is, he quotes a price at the tick $(N + k^*)T$ where $(N + k^*)T \geq a_s(q; c_s, I_s)$. If the order is filled at this quote, which is at least as large as his reservation price, his utility will increase. Q.E.D.

Proof of Proposition 3. For an order $q \in [q^*, q\dagger]$, for $q^* < q\dagger < q^{**}$, the equilibrium strategies of the specialist and the floor trader are the same as in proposition 1. The equilibrium strategy of the off-floor market maker as stated in lemma 3 is to quote $(N + k^*)T$ and offer a side payment to the broker. The broker's equilibrium strategy is to take the order to the off-floor market maker and receive the side payment as long as the posted quote of the specialist is not lower. Given these equilibrium strategies, the order flows to the off-floor market maker. Q.E.D.

Appendix B

An Example of Distortion When Customers Have Imperfect Information and Can Choose the Exchange on Which Their Order Is Executed

As we mentioned in note 20 above, we provide here an example in which we show that the practice of payment-for-order flow can result in inferior execution even if customers can choose the venue of order execution. The intuition we seek to illustrate is as follows. Suppose that customers have poorer information than brokers about the likelihood of getting better than quoted prices on the NYSE floor (a likely scenario). Then, if the broker can obtain payments for order flow off the exchange floor, he has an incentive to understate the probability of obtaining better than quoted prices on the exchange floor. This can result in situations in which the order is taken off-floor even though the broker knows that there is a high probability of obtaining superior execution on the exchange floor. Absent a regime with payment-for-order flow, the broker has no incentive to misrepresent priors, and the distortion described above can be eliminated.

The example to follow is highly structured, but we hope the analysis will convince the reader that the intuition is robust. Consider a fixed order size q and no asymmetric information. For simplicity, we will drop arguments from the ask price functions. Let N be the smallest integer such that $a_s < NT$. There are *two* floor traders, identical in every respect, and their inventories are such that $(N - 3)T < a_{f1} < (N - 2)T$ and $(N - 3)T < a_{f2} < (N - 2)T$; these floor traders compete in a Bertrand-like fashion for the order. The specialist and the broker know exactly the inventory and the carrying cost parameters of these floor traders (this assumption is made solely for simplicity).

There is a single off-floor market maker, and his inventory is such that $(N$

$-2)T < a_m < (N-1)T$. He can pay for order flow, specifically, he can make a onetime commitment to pay the broker an ϵ arbitrarily close to zero for the order. There is also a single customer, who can tell the broker where to execute the order. The customer observes the posted price on the NYSE and off-floor. Before the customer informs the broker of his decision, the customer asks the broker for his estimate of the probability of getting a better price than the minimum of the NYSE and off-floor quotes on the NYSE floor and then makes the decision.

Let us propose the following equilibrium prices: NT by the specialist and $(N-1)T$ by the off-floor market maker. We justify these equilibrium prices as follows.

If the off-floor market maker posts a price of NT or higher, the customer will tell the broker to take his order to the NYSE, which is not optimal. So the off-floor market maker posts $(N-1)T$.

The specialist's reservation price is such that he is unable to compete for the order. As he knows he will not receive the order in equilibrium, he is indifferent across all prices above his reservation price for the order. As in the text, we assume he picks the lowest price above his reservation price, NT.

The customer knows that the two floor traders are ex ante identical. Suppose the priors placed by the customer on the floor trader's reservation ask price are that $\mathrm{pr}(a_{fi} < [N-3)T] = 0$ and that

$$\mathrm{pr}[(N-2)T < a_{fi} < (N-1)T] = \phi_1$$

and

$$\mathrm{pr}[(N-3)T < a_{fi} < (N-2)T] = \phi_2.$$

We can specify these priors because the reservation ask price depends only on exogenous parameters.

The basis of the customer's choice between the exchanges will be the comparison of $\phi_1(N-1)T + \phi_2(N-2)T + (1-\phi_1-\phi_2)NT$ (the expected price on the NYSE floor) with $(N-1)T$ (the price quoted by the off-floor market maker). Suppose that ϕ_1 and ϕ_2 are sufficiently low that the condition

$$\phi_1(N-1)T + \phi_2(N-2)T + (1-\phi_1-\phi_2)NT > (N-1)T$$

holds.

With payment for order flow, the incentives of customers and brokers are not properly aligned. The broker has an incentive to lie about the probability of getting a price better than $(N-1)T$ on the NYSE floor (unity from his perspective) because he obtains a payment off the floor. For simplicity, suppose the customer's priors are common knowledge; in this case, one equilibrium is for the broker to tell the customer that there is a probability of exactly ϕ_2 of this happening and for the customer to tell the broker to execute the order off-floor (and get execution at $(N-1)T$).

Suppose that there is no payment for order flow. Since the broker has no incentive to lie in this case, we assume he tells the truth. In this case, the outcome will be different. The broker states that the probability of getting a price better than $(N-1)T$ on the NYSE floor is unity, the customer believes the broker and tells him to execute the order on the NYSE. The customer thus obtains a price of $(N - 2)T$, better than that obtained in a regime with payment-for-order flow.

References

Bernhardt, M., and Hughson, E. 1994. Discrete prices and interdealer competition. Working paper. Vancouver: University of British Columbia.

Biais, B., and Hillion, P. 1994. Insider and liquidity trading in the stock and options markets. *Review of Financial Studies* 7:743–80.

Christie, W., and Schultz, P. 1944. Why do NASDAQ market makers avoid odd-eighth quotes? *Journal of Finance* 49 (December): 1813–40.

Coffee, J. 1990. Brokers and bribery. *New York Law Journal* (September 27): 4.

Easley, D., and O'Hara, M. 1987. Price, trade size and information in securities markets. *Journal of Financial Economics* 19 (September): 69–90.

Glosten, L. 1991. Asymmetric information, the third market, and investor welfare. Working paper. New York: Columbia University.

Glosten, L., and Milgrom, P. 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics* 14 (March): 71–100.

Grossman, S., and Miller, M. 1988. Liquidity and market structure. *Journal of Finance* 40 (July): 617–37.

Harris, L. 1993. Minimum price variations, discrete bid/ask spreads, and quotation sizes. *Review of Financial Studies* 7:149–78.

Kyle, A. 1985. Continuous auctions and insider trading. *Econometrica* 53 (November): 1315–35.

Lee, C. 1993. Market integration and price execution for NYSE-listed securities. *Journal of Finance* 48 (July): 1009–38.

Lee, C., and Ready, M. 1991. Inferring trade direction from intraday data. *Journal of Finance* 46 (June): 733–46.

Madhavan, A. 1992. Trading mechanisms in securities markets. *Journal of Finance* 47 (June): 607–41.

Madhavan, A., and Smidt, S. 1991. A Bayesian model of intraday specialist pricing. *Journal of Financial Economics* 30 (November): 99–134.

McInish, T., and Wood, R. 1992. Hidden limit orders on the NYSE. Working paper. Memphis: Memphis State University.

New York Stock Exchange Constitution and Rules. 1992. New York: Commerce Clearing House.

Peake, J. 1992. Market center competition: Order flow or listings? Working paper. Englewood, N.J.: Peake/Ryerson Consulting Group Inc. Presented at the New York Stock Exchange/University of California, Los Angeles/University of Southern California Conference on Market Microstructure, Los Angeles, March 1992.

Wall Street Journal. 1991. SEC weighs switch to decimal stock trading (June 12).

Wall Street Journal. 1992. Big Board at age 200, scrambles to protect grip on stock market (May 13).

Wall Street Journal. 1994. Brokerage firms face investor suit alleging kickbacks (September 30).

Zabel, E. 1981. Competitive price adjustment without market clearing. *Econometrica* 49 (September): 1201–21.