Market Integration and Price Execution for NYSE-Listed Securities
Author(s): Charles M. C. Lee
Source: *The Journal of Finance*, Vol. 48, No. 3, Papers and Proceedings of the Fifty-Third Annual Meeting of the American Finance Association: Anaheim, California January 5-7, 1993 (Jul., 1993), pp. 1009-1038
Published by: Blackwell Publishing for the American Finance Association
Stable URL: http://www.jstor.org/stable/2329024
Accessed: 25/11/2009 11:33

THE JOURNAL OF FINANCE • VOL. XLVIII, NO. 3 • JULY 1993

Market Integration and Price Execution for NYSE-Listed Securities

CHARLES M. C. LEE*

ABSTRACT

For New York Stock Exchange (NYSE) listed securities, the price execution of seemingly comparable orders differs systematically by location. In general, executions at the Cincinnati, Midwest, and New York stock exchanges are most favorable to trade initiators, while executions at the National Association of Security Dealers (NASD) are least favorable. These intermarket price differences depend on trade size, with the smallest trades exhibiting the biggest per share price difference. Collectively, these results raise questions about the adequacy of the existing intermarket quote system (ITS), the broker's fiduciary responsibility for "best execution," and the propriety of order flow inducements.

IN THE EMERGING GLOBAL economy, the same security is often traded simultaneously at different physical locations. For such securities, market integration—the full and timely communication of intermarket information—is an issue of practical, academic, and regulatory importance.[1] In a fully integrated market, incoming buy (or sell) orders have an opportunity to be matched against the best available sell (or buy) orders across all locations. This intermarket matching process lowers the cost and time delay of trading and enhances the market's price efficiency. Conversely, in a poorly integrated (i.e., "fragmented") market, incoming orders may not be executed at the best available intermarket price. In such markets, the choice of where an order is routed can have a significant effect on the price obtained.

This paper explores the closely related issues of market integration and price execution for a sample of New York Stock Exchange (NYSE) listed securities. Most NYSE-listed securities also trade on at least one of five

* School of Business Administration, University of Michigan. I thank Roger Kormendi for directing me to this topic and the Mid America Institute for information and encouragement at the formative stage of the project. My academic colleagues, Vic Bernard, George Foster, Jim Hamilton, Larry Harris, Joel Hasbrouck, Marilyn Johnson, Mark Ready, Paul Seguin, Chester Spatt, Kent Womack, and Robert Wood all provided extremely helpful suggestions. I also benefited from the comments of Gene Finn (NASD), Bill Boye (Webco Securities), Joseph Mahoney (AC Partners), Bernie and Peter Madoff (Madoff Investments), Roger Hendrick (Midwest Stock Exchange), Chuck Black and Lee Korins (Pacific Stock Exchange), George Sofianos and Jim Shapiro (NYSE). Exceptionally detailed suggestions from an anonymous referee on an earlier draft greatly improved the paper. Any remaining errors or omissions are my own.

[1] The term "market integration" is used here to describe the extent that electronic linkages communicate the available trading opportunities at different physical locations. A fully integrated market is one in which all the price-relevant trading information available at each location is communicated quickly to the entire market.

regional exchanges and in the Over-The-Counter (OTC) market. These centers are linked by an electronic system (the Intermarket Trading System, or ITS) that immediately disseminates trades and quote revisions to all locations. In U.S. equity markets, exchange dealers must meet or beat the best ITS quote. Therefore, if the ITS system is fully communicating intermarket trading opportunities, identical market orders should have an equal opportunity for best price execution, regardless of initial routing locations. Conversely, if the ITS system does not fully communicate intermarket trading information, seemingly comparable orders may execute at different prices, depending on where they are routed and how they are handled at each location.[2]

The main proposition of this study is that the location of execution is price relevant for trades in NYSE-listed securities. The issue is timely, pertaining to the current regulatory debate surrounding payments for order flows. Since 1988, members of the National Association of Securities Dealers (NASD) and some regional specialists have paid brokers cash rebates, typically one cent per share, for directing customers orders to their market centers. These cash payments are one of many order flow arrangements between brokers and dealers (collectively called "order flow inducements") that may influence how brokers route their customers' orders.[3] Currently, the Securities and Exchange Commission (SEC) is considering the welfare implications of these inducements. One important issue under deliberation is whether the fiduciary responsibility of brokers to procure "best execution" for customers is compromised by side payments from competing dealers. Evidence of significant intermarket differences in the cost of trade execution would suggest possible conflicts of interest. Conversely, if price execution does not differ by trade location, these dealer-broker arrangements are unlikely to compromise the brokers' primary responsibility to their customers.

Using returns of matched stock portfolios, prior studies have investigated the effect of different market designs on asset prices (e.g., Reinganum (1990)). This study, however, uses Institute for the Study of Security Markets (ISSM) data on individual trades to compare the price execution across different market locations. Three separate tests are conducted using trades and quotes from 1988 and 1989. The first test compares the "liquidity premium" paid on off-Board trades to that paid on adjacent NYSE trades, where the liquidity premium is defined as the absolute difference between the trade price and the midpoint of the bid-ask spread. The second test classifies all trades as buys or sells using the Lee and Ready (1991) algorithm and compares the trade price of off-Board buys (or sells) to adjacent NYSE buys (or sells). The

[2] Note that market integration is a necessary, but not sufficient, condition for location indifference. Even in a fully integrated market, dealers may strategically ignore the intermarket information (where laws permit), or traders may gravitate to certain markets for nonpecuniary reasons (see Harris (1992) for a discussion of the economic forces that cause markets to integrate or fragment).

[3] See the report of the Order Flow Committee of NASD (1991), for a review of background information and institutional details. Coffee (1991) provides a good summary of legal issues.

third test examines the relative likelihood of price improvement over the quoted prices by documenting the frequency of inside-the-spread trading across the different exchanges. For all three tests, off-Board trades are matched against NYSE trades of similar size in the same security executed within two minutes. Collectively, these tests provide a framework for comparing execution costs across different markets.

The main result from these tests is that the execution price of similar adjacent trades can differ systematically depending on the location of execution. Even after controlling for the security traded, the trade size, and the time of execution, we still observe significant average execution price differences by location. Most of these differences are attributable to trading inside the best ITS quote, suggesting that the ITS does not fully reflect the available intermarket liquidity.[4] These findings provide empirical evidence about the consequences of market fragmentation. For market practitioners, the intermarket price differences also provide a benchmark for the broader problem of evaluating dealer services.

All three tests document a systematic relation between trade location and execution price. The first test shows liquidity premiums are typically lower at the Cincinnati, Midwest, and New York stock exchanges and higher at the NASD. Similarly, the second test shows investors tend to pay lower prices for buys and obtain higher prices for sells on the NYSE relative to adjacent off-Board trades. On average, the price for non-NYSE trades is 0.7 to 1 cent per share less favorable than that of adjacent NYSE trades. This amount is comparable to the typical cash rebates in payment for order flow agreements suggesting that, during 1988 and 1989, off-Board dealers paying cash inducements were able to recoup these payments in the form of higher effective spreads.[5] In aggregate, estimates of the total additional cost of off-Board executions are \$13 to \$18 million for 1988 and \$36 to \$47 million for 1989.

Further, these tests show that the relative price performance of market centers differs by trade size. In the small trades (100 to 400 shares), the NYSE offers the best execution. In the midsize trades (500 to 4900 shares), the Cincinnati and Midwest exchanges significantly outperform the NYSE in both years, while the Pacific exchange outperforms the NYSE for 1988. Relative price performance in the large trade class (5,000 shares or more) is mixed, with the Midwest, Pacific, and Instinet (which caters exclusively to an institutional clientele) all performing well. Only the NASD performs consistently worse than the NYSE in all size categories.

The third test, which is conditional on spread size, shows that price performance is closely related to the frequency of inside-the-spread exe-

[4] Harris (1990) defines liquidity in terms of the willingness of some traders (often, but not necessarily, dealers) to allow others to trade on demand. Intermarket liquidity refers to the aggregate liquidity available across all market locations.

[5] This comparison is intended only to help put the intermarket price differences into economic perspective. No causality between the intermarket prices and the existence of dealer-broker arrangements is intended. In fact, many regional dealers did not pay for order flow during this time period.

cutions. The market centers with the most frequent inside-the-spread executions—e.g., the NYSE, Midwest, and Cincinnati—also had the most favorable prices. Conversely, the market center least likely to improve prices over the prevailing ITS quote—the NASD—had the least favorable trade prices. Separate analyses show these results are not attributable to the tighter spreads on NASD trades. In fact, the frequency of price improvement on NASD trades is relatively low for all spread sizes.

Supplemental tests show that these findings are not specific to the choice of the two-minute window used for matching trades, since similar results obtain using one-, five-, or ten-minute windows. These findings are also not attributable to nonsynchronous reporting of trades and quotes on the ISSM tape. When quotes that are five seconds, two minutes, and five minutes ahead of each trade are used, similar intermarket rankings result. Moreover, these findings are not due to cross-sectional differences in firm characteristics, since similar results are obtained when execution costs are aggregated first by firm, then across firms.

Several important caveats should be kept in mind in interpreting these results. First, since the ISSM data present only trade prices, not orders, the identity of the trade initiator is unknown. This study assumes that all trades are initiated by outside public traders, who pay for liquidity. Consequently, the results pertain to execution costs for market orders, not limit or other price-contingent orders (unless these orders are executable upon receipt).[6] Further, trades that are not initiated by public orders (e.g., trades initiated by market makers) are not analyzed separately. Although the exact effect of these trades is difficult to quantify, they represent a relatively small percent of total trades and are not expected to alter significantly the findings reported here.[7]

Second, these tests focus on only one aspect of order execution—the trade price. While the average trade price provides a useful benchmark for price performance, the evaluation of dealer services is a broader problem. Factors such as the speed of execution, the amount of guaranteed depth (shares available at each price), and the reliability of trade settlement are all relevant in assessing execution quality. The brokers' fiduciary responsibility to procure "best execution" for their clients may involve tradeoffs along these different dimensions. Therefore, these results should be viewed as just one element in the evaluation of relative market center performance.

Third, while these results document an association between location and trade price, they do not establish causality. Causality seems to flow from location to price performance, but the reverse could also hold: that is, trades may be sent to particular locations in anticipation of price performance. For

[6] See Harris and Hasbrouck (1992) for a discussion of order submission strategies for market versus limit orders.

[7] Trades initiated by market makers are a subset of the intermarket orders sent across the ITS system. Except for the Cincinnati exchange, total trades from ITS orders represent at most 15 percent of the trades at each regional exchange. The related issue of regional trades arising from "limit book protection" is discussed in Section VII.

example, trades may be sent to a regional exchange only when it provides the best price. This possibility is particularly important for locations with relatively low trading volume, such as Instinet or Cincinnati. However, it does not explain the NASD results, since trades seem to be sent to there in spite of inferior price performance. The higher volume of the NASD in recent years may be attributable to payments for order flows, a practice the NASD endorses. However, it could also be explained by better dealer performance on the NASD in other dimensions unrelated to price execution.

Finally, we should not equate lower execution costs with superior market maker performance. In a pure dealer market, such as the NASD, the market maker is the sole supplier of liquidity. However, at the NYSE and most other exchanges, price improvements do not necessarily take place at the specialist's expense. At the NYSE, in particular, where specialist participation is estimated to be 13 to 22 percent of total share volume (Hasbrouck and Sofianos (1992) and Mann and Seijas (1991)), price improvement occurs largely at the expense of public limit orders.[8] The tests in this study measure relative costs of liquidity, but do not evalaute the contribution of specialists to the liquidity supply.

The remainder of the paper is organized as follows. Section I provides background on the order flow debate, Section II discusses why execution costs might differ across market centers, Section III describes the data, Sections IV through VI report the results of the three tests, Section VII addresses other methodological concerns and presents supplemental analyses, and Section VIII concludes.

I. Order Flow Inducements

The market for dealer services is highly competitive. A steady flow of customer orders allows the market maker to make a "dealer's turn"—the profit inherent in purchasing at the bid and selling at the ask.[9] Interestingly, the competition among dealers to attract order flow has focused on brokers,

[8] Price improvement can take place on an exchange floor in several ways. First, an incoming order can "cross" with a floor trader with a public order inside the spread (executed against the "crowd"). Alternatively, the specialist may elect to execute the order against his own account inside the quoted spread. Finally, the specialist can "stop" an order, guaranteeing execution at prices at least as favorable as the prevailing quotes. If an offsetting public order arrives in the next few minutes, the two are crossed inside the spread, resulting in price improvement for both sides. If no offsetting order appears, the stopped order is executed at the guaranteed price. In an extreme example, if the specialist only "stops" an incoming market order when the quotes represent public limit orders, then all price improvement from a midpoint cross occurs at a cost to public limit orders.

[9] Dealers also profit from order flows through the division of revenue from the Consolidated Tape Association (CTA). Brokers and other market participants pay CTA for the dissemination of trade information. This revenue is divided among the market centers on the basis of the number of transactions each reports.

rather than directly on investors. Since most investors do not specify a preference for trade location, this decision is typically made by the brokers.

To attract orders, competing market centers offer brokers a myriad of inducements (see NASD (1991)). The most controversial of these inducements involves cash payments. In a typical payment for order flow arrangement, the market maker agrees to pay the broker one cent per share for all directed market orders. The dealer also establishes certain conditions to be met for cash payments. For example, NASD (1991) cites these common stipulations in payment for order flow agreements: trading spreads of at least one eighth, minimum monthly order flow (usually a minimum of 100,000 shares), maximum shares per order (usually 3,000 shares), order source restrictions (no professional or program orders), and no payments for limit orders or orders for low-priced stocks.

Neither brokers nor dealers are required to report the total amount of their payments for order flow.[10] Although estimates of the extent of these activities vary widely (Steptoe (1989), *Securities Week* (1989a, 1989b)), the general consensus is that the practice started in earnest in 1988 and has grown rapidly since. Payment for order flow arrangements are most often associated with the NASD, where the practice first started, with participation from the regional specialists increasing since 1989 (Crawford (1991), NASD (1991)).

One key issue for the SEC is whether customer orders entrusted to a broker should be used as a bargaining chip in contract negotiations. Proponents of payment for order flow inducements argue that investors are not harmed by these arrangements. They note that payment for order flows is in the spirit of fair competition, which in turn lowers execution costs. Moreover, such cash payments are no different from other "soft dollar" arrangements between brokers and dealers, where order flows are exchanged for reciprocal services. Opponents of payments for order flow claim that such payments exacerbate the selective market making of regional dealers (i.e., "cream skimming"). Unlike NYSE specialists, who must make a market for all NYSE-listed stocks, purchasers of order flows can target the more profitable "low end" business, which consists mainly of small trades in more liquid stocks. The NYSE specialist, finding profit margins reduced, may look to recover these losses by increasing the quoted spreads and, thus, the overall liquidity costs.[11]

[10] Brokers provide a standard notification to customers indicating that the firm may have received additional remunerations in connection with a customer's order. This notice typically appears on the back of confirmation slips.

[11] Glosten (1991) provides a formal treatment of the economic tradeoff between lower spreads from increased competition and higher spreads due to cream skimming. Empirical tests of the Glosten model using bid-ask spreads are complicated by the possibility that NYSE specialists may widen spreads (and increase inside-the-spread trading on the floor) for strategic reasons, to deter nonprimary exchange traders from free riding on the NYSE price discovery mechanism. This type of gamesmanship is difficult to distinguish from a spread effect due to "cream skimming."

Implicit throughout the regulatory debate is the assumption that the market for NYSE-listed securities is fully integrated, so price execution is the same at all locations. With market integration, the broker's responsibility to procure best execution is unlikely to be compromised by side payments from market makers. Therefore, if an examination reveals that execution costs are not significantly different by location, the debate should focus primarily on the effects of cream skimming versus increased competition. However, if execution costs differ by location, broker-dealer independence (both perceived and substantive) may be of real concern. In particular, if market centers that pay for order flow offer significantly worse price execution, the propriety of order flow inducements may need to be reexamined.

II. Execution Costs and Trade Location

Several reasons lead us to expect a relation between execution costs and trade location. First, and perhaps most important, the ITS quote does not fully reflect available intermarket liquidity. While all the market makers guarantee execution at the best intermarket price, as displayed on the ITS, much of the trading actually takes place inside the ITS spread (e.g., Lee and Ready (1991)). This activity reflects "hidden liquidity" at local markets that is not readily communicated by electronic links such as the ITS. In the case of the NYSE or a regional exchange, this liquidity may be provided by floor traders or the specialist—parties who may not wish to advertise formally their proclivity to trade.[12] The relative performance of individual market centers depends on their ability either to attract this liquidity, or provide it in a dealer capacity (see Harris (1990)). Since each market center has its own procedure for handling customer orders and attracting liquidity suppliers, the average cost of execution can differ by location.

The NYSE has long argued that the most favorable execution occurs on the primary exchange floor, where inside-the-spread executions occur frequently as specialists and floor brokers compete to supply liquidity. In response to this argument, regional specialists and NASD dealers point out that the spread for much of the order flow executed at their centers is only one eighth, leaving little room for price improvement (Stern (1989)). When the spread is wider, regional executions also transact frequently inside the best ITS quote. In fact, most regional exchanges as well as some NASD dealers have developed their own algorithms to improve the likelihood of inside-the-spread executions (see Appendix A for summary). The relative effectiveness of these

[12] The NASD operates under a multidealer electronic quote system, without an exchange floor, so floor trading does not explain inside-the-spread trades at the NASD. However, NASD dealers can, and do, improve prices relative to the ITS spread by facilitating execution inside the best ITS quote. This type of liquidity is also "hidden," in the sense that it is not reflected in the ITS system. Mann and Seijas (1991) provide a nontechnical discussion of differences between the NYSE specialist system and the NASD dealer system.

algorithms in reducing execution costs for public market orders is an open empirical issue.

Few prior studies have examined trade execution performance across markets. Reinganum (1990) finds differences in the monthly returns of NYSE and NASD firms matched on size and risk characteristics. He suggests that the multiple dealer market of the NASD may have a liquidity advantage for small firms, but not for large firms. Mayer and Leigh (1991) find that daily extreme prices (highs and lows) are more likely to be off-Board trades. They note that if one assumes the daily highs (or lows) tend to be buys (or sells), this finding is consistent with worse off-Board execution. These studies do not measure the liquidity premium per trade, or control for trade characteristics such as size and time of execution.

Using intraday data and different empirical methods, Blume and Goldstein (1991) provide an analysis much closer to this study. They also compute a liquidity premium per trade based on the absolute difference between the trade price and the quote "midspread." In their study, the average difference in liquidity premiums between NYSE and non-NYSE trades during 1989 is estimated to be 0.79 cents per share. However, Blume and Goldstein do not match regional and NYSE trades by their time of execution, nor do they report results for individual exchanges, or directly compare prices for buys and sells. Instead, their study makes extensive efforts to identify the best intermarket quote and provides detailed comparisons of NYSE and non-NYSE quote prices.

III. Data Description

The data used in this study are obtained from the Institute for the Study of Security Markets (ISSM). The ISSM data contain a record of all trades and quotes for NYSE and AMEX-listed firms. Each trade and quote is time stamped to the nearest second, with the originating exchange identified. The study period covers all 505 trading days in 1988 and 1989, except October 25, 1989. The quotes for this date are missing in the version of the ISSM tape used for this study, so all trades from this day are excluded. Opening batch trades on the NYSE (the first trade of each day not preceded by an opening quote) are excluded because they do not have a regional exchange counterpart. In addition, all late trades, trades reported out of sequence, or trades with special settlement conditions are excluded since their prices are not comparable to adjacent trades.

The sample of firms used is the first 500 NYSE common stocks on the 1988 and 1989 tape that were listed for the full year and had a CUSIP number. The firms are sorted by ticker symbol on the ISSM tape. Since it is unlikely that symbol order is related to execution costs, this sample should represent the total population of NYSE-listed securities. Table I reports the distribution of trades for the 500 sample firms in each year. Just over 7.3 million trades were transacted for these firms during 1988, of which 1.9 million (26 percent)

Table I

Distribution of Total Trades by Size and Exchange

Table values represent the number of trades (in thousands) in each exchange and size category for the 500 sample NYSE firms during 1988 and 1989.

	Trade Size (in Shares)						
	100–400	500–900	1000–1900	2000–4900	5000–9900	10,000 +	All
			Panel A: 1988				
NYSE	2409.1	977.8	896.2	653.8	288.6	229.1	5454.6
Boston	81.3	25.1	19.5	8.5	2.9	2.4	139.8
Cincinnati	6.7	9.1	10.1	6.2	1.2	1.2	34.6
Midwest	461.0	124.4	89.0	33.5	10.7	11.9	730.5
Pacific	363.5	135.1	81.7	24.0	4.4	2.4	611.0
Philadelphia	147.3	35.3	19.1	6.4	2.3	2.8	213.2
NASD	106.2	23.3	19.0	6.7	2.0	3.5	160.6
Instinet	0.3	0.3	0.5	0.6	0.4	0.5	2.5
Total	3575.4	1330.3	1135.1	739.6	312.3	253.9	7346.7
Non-NYSE	1165.7	352.5	238.9	85.8	23.7	24.8	1892.1
(%)	(32.9)	(26.5)	(21.0)	(11.6)	(7.6)	(9.8)	(25.8)
			Panel B: 1989				
NYSE	2694.4	1056.9	986.6	742.2	336.0	275.0	6091.1
Boston	157.2	41.4	30.7	11.9	4.4	3.1	248.7
Cincinnati	6.6	14.9	15.9	7.5	0.6	1.2	46.8
Midwest	594.6	156.4	116.9	46.4	12.4	13.0	939.6
Pacific	397.4	137.2	82.9	26.6	4.8	2.7	651.6
Philadelphia	185.7	38.6	20.9	7.5	2.9	3.5	259.0
NASD	260.2	49.1	36.6	12.1	2.7	4.5	365.3
Instinet	0.2	0.3	0.5	0.7	0.4	0.6	2.7
Total	4296.4	1494.8	1291.0	854.9	364.2	303.6	8604.9
Non-NYSE	1602.0	437.9	304.4	112.7	28.2	28.6	2513.8
(%)	(37.3)	(29.3)	(23.6)	(13.2)	(7.7)	(9.4)	(29.2)

were executed away from the NYSE. Similarly, 8.6 million trades were transacted in 1989, of which 2.5 million (29 percent) were executed off-Board. The proportion of regional trades for this sample is similar to that reported for the total population of NYSE-listed securities during this period. The overall movement of volume away from the NYSE, with the NASD as the main beneficiary, is also consistent with the aggregate statistics reported in the NYSE *Fact Book* (1991).

With the exception of the Cincinnati exchange and Instinet, the regional exchanges tend to execute smaller trades. For example, in 1988 trades of 900 shares or less (a rough proxy for individual investor trades) represent over 75 percent of all transactions in NYSE-listed securities on the Boston, Midwest, Pacific, and Philadelphia exchanges, as well as the NASD. On the NYSE, only 62 percent of the total trades are 900 shares or less. These differences may be due, in part, to the maximum order size stipulation of many order flow arrangements.

IV. Liquidity Premium Tests

A. Full Sample

In the first set of tests, the liquidity premium for each trade is computed as the absolute difference between the actual trade price and the "midspread" (average of bid and ask prices) of the prevailing quote at the time of the trade. Assuming that, on average, the specialist's spread is set symmetrically around the equilibrium price, the liquidity premium provides an estimate of the effective "half-spread" for each trade.[13] The liquidity premium paid on off-Board trades is compared to the liquidity premium paid on NYSE trades. All quotes eligible for the Best-Bid-Or-Offer (BBO) calculation, except regional exchange "autoquotes," are used.[14]

Case 1 of Figure 1 illustrates the liquidity premium test. Regional trades T4 and T6 have liquidity premiums of 12.5 cents and 0 cents per share, respectively. Assuming the NYSE trades T3 and T5 are matched to T4, then the excess cost of the T4 execution is computed as 6.25 cents per share in favor of the NYSE. Similarly, the excess cost for T6 is -6.25 cents per share, favoring the regional exchange.

Table II provides a profile of the average liquidity premium paid for the trades in each size and exchange category, expressed in cents per share. For 1988, the average liquidity premium on NYSE trades is 9.6 cents per share, which is slightly lower than the average liquidity premium on off-Board trades. However, intermarket comparisons based on these numbers may be misleading for two reasons. First, if regional dealers "skim the cream" by making markets only in more liquid stocks, regional trades should have a lower average liquidity premium than NYSE trades. Thus, lower execution costs may, in fact, be a product of the stock selection procedure followed by regional dealers. Second, quoted and effective spreads are known to display pronounced intraday patterns (e.g., Lee, Mucklow and Ready (1993)). The exact reasons for these patterns are not known, but their existence suggests it may not be appropriate to match trades from different times of the day. To

[13] Following the recommendation in Lee and Ready (1991), any quote in the five seconds preceding the trade is ignored in favor of the previous quote. This mitigates against situations where a quote triggered by a trade is recorded on the ISSM tape with the same time stamp as, or just ahead of, the triggering trade.

[14] The current version of the ISSM tape does not contain noncompetitive regional quotes (defined as regional quotes which are worse than NYSE quotes). Most of the excluded quotes are electronically generated "autoquotes" of 100 shares each. Such quotes automatically adjust as NYSE quotes change to avoid becoming the best available quote. In theory, the best ITS bid or offer cannot be identified without a full set of regional quotes, because a regional quote that is noncompetitive when issue may become competitive when the prevailing quote changes. However, as discussed later, the qualitative results remain the same when only NYSE quotes are used.

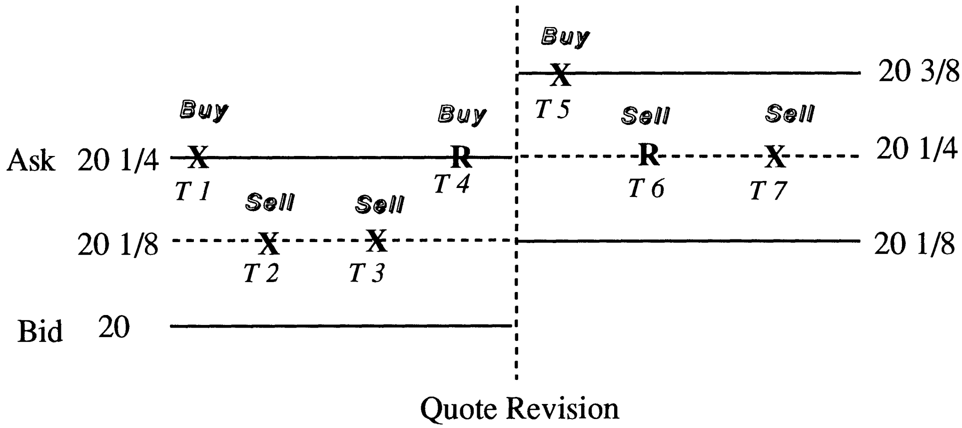

Figure 1. Liquidity premium and directional price tests. NYSE trade (X). Regional or NASD trade (R).

Case 1: Liquidity Premium Test

The liquidity premium (LPrem) for each trade is defined as the distance between the trade price and the midspread of the best ITS quote at the time of the trade. The Liquidity Premium Test compares the LPrem for each regional trade to the LPrem for adjacent NYSE trades.

Regional Trade	Liquidity Premium on Regional Trade (LPremReg)	Matching NYSE Trade(s)	Liquidity Premium on NYSE trade(s) (LPRemNYSE)	Excess Cost of Regional Execution (LPRemReg − LPremNYSE)
T4	12.5 ¢/share	T3, T5	(0 + 12.5)/2 = 6.25 ¢/share	6.25 ¢/share (favors NYSE)
T6	0 ¢/share	T5, T7	(12.5 + 0)/2 = 6.25 ¢/share	−6.25 ¢/share (favors regional)

Case 2: Direction Price Test

Each trade is classified as a buy or sell, using the best ITS quote at the time of the trade and the price of adjacent trades. The price of regional buys (or sells) is then compared to the price of adjacent NYSE buy (or sells).

Regional Trade	Direction of Trade	Regional Trade Price (PrReg)	Matching NYSE Buy of Sell	NYSE Trade Price (PrNYSE)	Excess Cost of Regional Trade (PrReg − PrNYSE)
T4	Buy	20 1/4	T5	20 3/8	−12.5 ¢/share (favors regional)
T6	Sell	20 1/4	T7	20 1/4	0

Table II

Average Liquidity Premium by Size and Exchange

Table values represent the average liquidity premium paid for the trades in each size and exchange category, expressed in cents per share. The liquidity premium is computed as the absolute difference between the trade price and the midspread of the best ITS quote at the time of the execution. The midspread is the average of the bid price and ask price.

	Trade Size (in Shares)						
	100–400	500–900	1000–1900	2000–4900	5000–9900	10,000 +	All
Panel A: 1988							
NYSE	9.2	9.6	9.8	10.0	10.0	10.5	9.6
Boston	11.1	11.7	11.1	11.2	11.3	11.2	11.2
Cincinnati	13.1	12.0	9.7	12.1	9.9	9.5	11.4
Midwest	10.1	9.6	9.0	8.6	9.1	9.1	9.8
Pacific	10.4	10.1	9.1	7.9	7.3	8.5	10.0
Philadelphia	10.4	10.6	9.6	9.5	10.4	9.9	10.4
NASD	9.7	9.6	9.4	9.8	9.9	10.6	9.7
Instinet	11.1	9.7	9.6	9.9	9.2	9.1	9.7
All	9.5	9.8	9.7	9.9	9.9	10.4	9.7
Panel B: 1989							
NYSE	8.5	8.7	8.7	8.7	8.8	15.6	9.2
Boston	9.6	9.6	9.4	9.6	9.8	10.2	9.6
Cincinnati	13.6	11.9	9.0	11.0	9.3	8.5	10.9
Midwest	9.2	8.6	8.2	8.1	8.1	8.4	8.9
Pacific	9.4	9.2	8.5	7.5	7.1	8.2	9.1
Philadelphia	10.7	9.8	9.4	9.0	9.1	9.4	10.3
NASD	9.5	9.0	8.8	8.7	9.1	9.3	9.3
Instinet	10.3	9.1	8.6	8.9	8.0	7.5	8.6
All	8.9	8.8	8.7	8.7	8.8	15.4	9.2

control for this potential selection bias, this study uses a time-matched design in all the following tests.[15]

B. Time-Matched Sample

Tables III and IV control for the potential selection bias created by cream skimming. Table III shows the distribution of regional trades that have at

[15] The findings in Table II provide only weak evidence for the prediction that larger trades involve higher liquidity premiums (e.g., Easley and O'Hara (1987)). This may be due to the pooling of different firms in the analysis, since firms with greater frequency of large trades are likely to have lower average liquidity premiums. Alternatively, the larger trades tend to be brokered nonanonymously, thus reducing the adverse selection costs. The large average liquidity premium for 1989 NYSE trades over 10,000 shares is due to outliers that do not affect the time-matched sample, since they do not have matching regional trades.

Table III

Distribution of Off-Board Trades in Matched Sample

Table values represent the number of regional or OTC trades (in thousands) with at least one adjacent trade of similar size on the NYSE. NYSE trades are considered adjacent if they are in the same size class and occur in the same security, on the same day, and within two minutes of the regional or OTC trade.

	Trade Size (in Shares)						
	100–400	500–900	1000–1900	2000–4900	5000–9900	10,000 +	All
			Panel A: 1988				
Boston	32.6	6.1	5.3	2.3	0.6	0.4	47.3
Cincinnati	3.2	3.2	4.2	3.0	0.3	0.2	14.2
Midwest	184.3	30.1	24.5	9.1	1.7	1.7	251.5
Pacific	148.6	32.1	21.7	5.8	0.7	0.4	209.3
Philadelphia	59.7	8.3	4.9	1.5	0.3	0.4	75.0
NASD	54.3	6.7	6.0	1.8	0.3	0.4	69.4
Instinet	0.1	0.1	0.1	0.1	0.0	0.0	0.5
Total	482.6	86.6	66.6	23.7	3.9	3.5	667.2
			Panel B: 1989				
Boston	69.2	10.9	8.9	3.5	0.9	0.6	94.1
Cincinnati	3.3	5.4	6.4	3.6	0.1	0.2	19.2
Midwest	237.9	39.1	33.5	13.4	2.4	2.3	328.6
Pacific	174.0	34.3	22.9	6.8	0.8	0.5	239.3
Philadelphia	80.9	10.1	6.2	2.0	0.5	0.6	100.3
NASD	136.8	15.1	11.9	3.3	0.4	0.5	167.9
Instinet	0.1	0.1	0.1	0.2	0.1	0.1	0.6
Total	702.2	115.0	90.1	32.8	5.2	4.7	950.0

least one matching trade on the NYSE. A NYSE trade is considered a match if it is for the same security, in the same size class, and occurs on the same day within two minutes of the regional trade. A later test examines results using matching windows of one, five, and ten minutes. Using a two-minute window, approximately 667,000 (35 percent) of the off-Board trades in 1988 have at least one adjacent trade of similar size on the NYSE. Similarly, in 1989, approximately 950,000 (38 percent) off-Board trades have a matched observation on the NYSE. If more than one matching NYSE trade is found, the liquidity premium for the control group is the average of all qualifying NYSE trades. All regional trades executed between 9:40 A.M. and 3:50 P.M. are eligible for inclusion in this sample.[16] A total of 64 paired observations (5 in

[16] The exclusion of regional trades in the first and last ten minutes of trading ensures the distribution of matching NYSE trades has an equal chance of occurring in any of the ten minutes on either side of a regional trade. It also mitigates against the strong market trends at the opening of trading reported by Wood, McInish, and Ord (1985). However, none of the main results are changed if regional trades in the first and last ten minutes are included.

Table IV

Average Excess Liquidity Premium Paid for Off-Board Trades in Matched Sample

Table values represent the average excess liquidity premium paid for off-Board trades, expressed in cents per share (negative values represent savings relative to the NYSE execution). The liquidity premiums on off-Board trades are compared to the liquidity premiums for adjacent NYSE trades of similar size. NYSE trades are considered adjacent if they occur within two minutes of the regional or NASD trade. The average by trades is the excess liquidity premium paid for the average trade at that exchange; the average by shares is the excess liquidity premium paid for the average share traded at that exchange (each trade is weighted by the number of shares transacted).

	Trade Size (in Shares)						Average	Average
	100–400	500–900	1000–1900	2000–4900	5000–9900	10,000+	by Trades	by Shares
				Panel A: 1988				
Boston	1.4*	0.71	0.67*	−0.03	2.11*	1.83*	1.23*	1.20
Cincinnati	1.36*	−1.21**	−1.36**	−0.02	0.73	1.54	−0.34**	0.43
Midwest	0.68*	−0.70**	−0.95**	−0.90**	−0.28**	−0.69**	0.28*	−0.20
Pacific	1.10*	−0.23**	−0.18**	−0.50**	−0.71**	−1.47**	0.71*	0.01
Philadelphia	1.56*	0.50*	−0.50	−0.41	0.32	−0.51	1.26*	0.35
NASD	1.51*	0.76*	0.59*	0.21*	2.61*	0.31*	1.32*	1.10
Instinet	2.21	0.67	−0.46	−0.52	0.48	−3.93	0.05*	−4.02
All	1.07*	−0.22**	−0.42**	−0.49**	0.31	−0.28	0.69*	0.20
				Panel B: 1989				
Boston	1.04*	0.55*	0.33	0.35	0.88*	1.61*	0.89*	0.91
Cincinnati	2.65*	0.44	−0.44**	1.49*	1.32	1.57	0.74	1.62
Midwest	0.87*	−0.03**	−0.27**	−0.32**	−0.06	−0.25	0.58*	−0.01
Pacific	1.19*	0.38	0.23	−0.02**	0.42	0.58	0.94*	0.48
Philadelphia	1.82*	1.21*	0.89*	0.46	1.26	1.17*	1.66*	1.28
NASD	1.58*	1.23*	0.89*	1.13*	2.26*	1.00*	1.49*	1.29
Instinet	1.72	1.42	1.81*	1.35	0.77	0.07	1.33*	0.81
All	1.22*	0.45*	0.14	0.22**	0.51	0.46	0.98*	0.59

* Favors the NYSE and is significantly different from zero at 0.01 level (two-tailed) using the Fisher sign test.

** Favors the regional exchange and is significantly different from zero at 0.01 level (two-tailed) using the Fisher sign test.

1988, 59 in 1989) are removed as a result of outlier tests. The main filter is the requirement that the absolute difference in liquidity premium between adjacent NYSE and regional trades be less than $2.00 per share.

Table IV reports the excess liquidity premium paid on the regional exchange relative to the liquidity premium for the control group of matched NYSE trades. Table values represent the liquidity premium for the regional trade minus the liquidity premium for the paired NYSE trade(s), averaged across all pairs and expressed in cents per share. Positive (or negative) values reflect higher (or lower) execution costs on the regionals. To average across trade sizes, two calculations are made for each exchange:

1. Average by trades

$$\text{Average Excess Liquidity Premium} = \frac{\sum_{i=1}^{n}(\text{LPremReg}_i - \text{LPremNYSE}_i)}{n}$$

2. Average by shares

$$\text{Average Excess Liquidity Premium} = \frac{\sum_{i=1}^{n}(\text{LPremReg}_i - \text{LPremNYSE}_i) \times \text{Shr}_i}{\sum_{i=1}^{n}\text{Shr}_i}$$

where:

n	= the number of matched observations for this exchange.
LPremReg_i	= the liquidity premium paid on the regional trade in the ith matched sample for this exchange
LPremNYSE_i	= the liquidity premium paid on the NYSE trade(s) in the ith matched sample for this exchange
Shr_i	= the number of shares traded for the regional trade in the ith matched sample for this exchange.

The first measure reflects the additional liquidity premium paid on the average trade executed at the given exchange. The second measure places greater weight on the larger trades and captures the additional liquidity premium paid on the average share executed at the given exchange.

The number of positive values in both panels of Table IV shows the liquidity premium is generally lower for NYSE trades. The NYSE advantage is most pronounced in the 100 to 400 shares class. The regional markets receive a disproportionately large share of these trades, yet appear to provide poorer price execution. For small trades, a Fisher sign test of statistical significance readily rejects the null hypothesis that the execution cost of off-Board trades is the same as the execution cost of adjacent NYSE trades. Given the large number of observations, this result is not surprising. Perhaps more important, however, is the economic magnitude of the estimated differences. The average excess liquidity premium per trade (the next to last column of Table IV) shows that off-Board trades generally involve higher execution costs in the order of 0.5 to 1.5 cents per share. The equally weighted average excess liquidity premium per non-NYSE trade is 0.69 cents per share in 1988 and 0.98 cents per share for 1989. Interestingly, this is approximately the magnitude of the payments in many order flow agreements. The evidence suggests that the higher effective spreads in the off-Board trades is sufficient to cover the cost of these payments.

The last column shows the volume-weighted difference by exchange. Here the difference between NYSE and non-NYSE performance is less pronounced, reflecting the better performance of the regional exchanges in the larger trade classes. In 1988, the Midwest exchange and Instinet had better volume-weighted performance than the NYSE. In 1989, only the Midwest performed better.

Although price execution is generally more favorable at the NYSE, results vary significantly by trade size and location. For example, the Cincinnati, Midwest, and Pacific exchanges tend to perform better than the NYSE in the midsized (500 to 5,000 share) trades. The NASD trades, however, display consistent higher liquidity premiums than their matching NYSE trades. This result holds in each size category in both years. Instinet, an electronic display system that allows large financial institutions to trade directly with each other, tends to perform better than the NYSE in the larger trade sizes.

C. Explaining the NASD Results

One of the most consistent results in Table IV is the performance of the NASD. In most trade sizes and across both years, NASD executions seem the least favorable to investors. These results may reflect the greater reliance of NASD dealers on intermarket quotes. Section VI examines this possibility by evaluating the frequency of inside-the-spread trades. The NASD result may also reflect the method of measuring liquidity premium. An important feature of the liquidity premium test is its reliance on the quote midspread at the time of the trade. This computation does not include any price "slippage" due to a movement in the quoted spread between the time of the order issuance and trade execution.[17] For example, if the midspread moves higher by one eighth between the time a buy order is issued and when it is executed, the additional one-eighth cost will not be included in the liquidity premium computed above. In short, this approach measures only the execution cost relative to an existing spread, when an investor may, in fact, face a significant additional risk due to slippage. The potential effect of price slippage is examined next.

V. Tests Based on Buy and Sell Prices

A. Directional Price Tests

Some argue that best price execution is more than simply execution inside the bid-ask spread (e.g., NASD (1991)). During 1988 and 1989, NASD dealers

[17] Price slippage is a natural consequence of adverse selection costs as market makers adjust prices to reflect the information revealed in incoming trades. See Glosten and Milgrom (1985) or Easley and O'Hara (1987) for formal treatments. Empirically, Hasbrouck (1988), Lee and Ready (1991), and Petersen and Umlauf (1991) show that buys (or sells) tend to be followed by upward (or downward) revisions in the specialist's quote. Also see Perold (1988) for an alternative way to measure slippage costs. Perold's "implementation shortfall" approach compares the actual profit from a real portfolio to a paper portfolio. It cannot be used here since no information is available about an order's time of origination.

typically executed at the best ITS bid-ask quote without attempting further price improvement.[18] Although this strategy reduces the chance of an inside-the-spread trade, it does provide faster execution. Since the benefit of an inside-the-spread execution may be outweighed by the cost of price slippage, NASD dealers may provide superior price execution that the liquidity premium test fails to capture. The tests in this section directly compare the trade price for buys (or sells) on regional exchanges to the trade price for adjacent buys (or sells) on the NYSE. If the price advantage from faster executions is greater than the price advantage of inside-the-spread executions, the NASD performance should improve under these tests.

Lee and Ready (1991) propose an algorithm that classifies each trade as buyer or seller initiated. This algorithm (summarized in Appendix B) relies on the prevailing bid and ask prices as well as the prior price changes ("tick tests") in classifying trades. In this study, the algorithm provides a direct way of comparing the price of buys (or sells) executed in a regional exchange to the price of adjacent buys (or sells) on the NYSE. The market that provides better execution should have lower average trade prices on market buys and higher average trade prices on market sells. The same algorithm is applied to trades from all exchanges, so even though trade misclassifications may introduce noise, they should not bias the results in favor of a particular exchange.

Case 2 of Figure 1 illustrates the directional trade test. The key trade in this sequence is T4. Under the liquidity premium test, this trade is compared to T3 and T5, and the regional execution is less favorable by 6.25 cents. However, the quote revision between T4 and T5 (i.e., price slippage) shows that the initiator of T4 actually bought at a cheaper price than the initiator of T5. The directional price test recognizes the better price obtained on the T4 trade and credits the regional exchange for a more favorable execution.[19]

Table V reports the average difference between the price of buyer-initiated regional trades and the price of matching buyer-initiated trades on the NYSE, expressed in cents per share. Positive table values indicate a higher price paid for buys on the regional exchanges. Negative values indicate savings from the NYSE. The number of positive values in Table V suggests that the price obtained for similar buys is generally higher for regional executions. These findings are consistent with the results of the liquidity premium tests in Table IV. Once again, the NYSE advantage is most evident

[18] More recently, some regional and OTC market makers have established methods to improve price execution by exposing an investor's order inside the quoted spread for a limited period of time. A primary example of this is Madoff Investment's MISSION system, implemented in the latter part of 1990. These price improvement algorithms may improve NASD and regional test results for future studies using post-1989 data.

[19] T4 is classified as a buy in the Lee and Ready (1991) algorithm because it is executed at the ask price. Note that it is also a buy using the "tick test" (based on the direction of the previous price change). It would be classified as a sell using the "reverse tick test" (based on the direction of the next price change), but Lee and Ready (1991) show the reverse tick test is generally inferior to the tick test.

Table V

Average Additional Cost for Off-Board Buys

Table values represent the average additional cost of buying for off-Board trades, expressed in cents per share (negative values represent savings relative to the NYSE execution). The Lee and Ready (1991) algorithm is used to infer buy-sell directions. The price of each off-Board buy is compared to the price for two adjacent NYSE buys of similar size. NYSE buys are considered adjacent if they occur within two minutes of the off-Board trade. The additional cost of buying is the difference between the off-Board trade price and the NYSE trade price (PrReg − PrNYSE). Sample size is 207,818 matched buy executions for 1988 and 332,121 matched buy executions for 1989.

	Trade Size (in Shares)						Average by Trades	Average by Shares
	100–400	500–900	1000–1900	2000–4900	5000–9900	10,000+		
Panel A: 1988								
Boston	1.14*	0.35	0.07	−0.02	0.15	2.74	0.84*	0.70
Cincinnati	0.60*	−0.71**	−0.82**	0.01	0.95*	0.51	−0.29**	−0.17
Midwest	0.25°	−0.69**	−0.56**	−0.93**	−0.12	−0.69**	−0.04	−0.39
Pacific	0.79*	−0.69**	−0.42**	−0.57**	−0.91**	−0.06	0.36*	−0.21
Philadelphia	1.41*	0.41*	−0.27**	0.10	0.17	0.56	1.12*	0.46
NASD	1.58*	1.23*	0.85*	0.41	0.21*	1.25	1.44*	0.65
Instinet	−2.84	0.00	1.07	0.33	1.41	−0.83	−0.03	0.01
All	0.78*	−0.38**	−0.35**	−0.46**	−0.07	0.13	0.43*	−0.04
Panel B: 1989								
Boston	0.65*	0.13	0.15	−0.01	0.61	0.65	0.52*	0.30
Cincinnati	0.32	−0.71**	−0.49**	0.58*	0.30	0.90	−0.18**	0.26
Midwest	0.66*	−0.32**	−0.41**	−0.43**	−0.19	−0.34	0.35*	−0.18
Pacific	0.80*	−0.55**	−0.36**	−0.21**	0.04	0.51	0.45*	0.25
Philadelphia	1.64*	0.49*	0.29	0.41	1.10	1.26*	1.40*	0.97
NASD	1.59*	1.15*	1.01*	1.02*	2.85*	0.74	1.50*	0.77
Instinet	−1.32	0.31	2.89	1.06	1.73	0.00	1.03*	1.03
All	0.99*	−0.12**	−0.12**	−0.02**	0.29	0.21	0.70*	0.23

* Favors the NYSE and is significantly different from zero at 0.01 level (two-tailed) using the Fisher sign test.

** Favors the regional exchange and is significantly different from zero at 0.01 level (two-tailed) using the Fisher sign test.

in the small trades, with the Cincinnati, Midwest, and Pacific exchanges doing well in the midsize trades. The results for seller-initiated trades are symmetrical and also favor the NYSE; prices for NYSE sells are generally higher than prices for regional sells. The magnitudes of the price differences for seller-initiated trades (not reported) are similar to those reported in Table IV for buyer-initiated trades.

These results show that price slippage is not a significant factor in explaining the NASD performance. For example, the average excess cost per trade for 1988 using the directional price test is 0.63 cents per share (0.43 for buys and -0.84 for sells), which is only slightly lower than the 0.69 cents per share reported in Table IV. Under the directional price test, the average excess cost for NASD actually increases from 1.10 cents per share to 1.45 cents per share. For the NASD trades in this sample, faster execution does not translate into lower execution costs.

B. An Estimate of Aggregate Costs

Table VI presents an estimate of the aggregate additional costs borne by investors whose orders are executed away from the NYSE. The per share execution costs are based on the time-matched sample, which includes over one third of the total off-Board trades. In Table VI, the volume-weighted per share cost is applied to the total share volume in each exchange to obtain an estimate of the total additional cost of regional trades. Panel A provides an estimate based on the excess liquidity premium method used to produce Table IV. Panel B provides an estimate based on the average additional cost using the directional price tests. These costs are obtained by averaging additional regional costs on all matched observations of buys and sells.

For 1988, the estimated additional cost of execution borne by investors is \$13 million using the liquidity premium approach and \$18 million using direct buy-sell prices. These totals increase to \$47 million and \$36 million respectively in 1989. The increase from 1988 to 1989 is due not only to a volume shift away from the NYSE, but also to an increase in the additional cost of off-Board execution. Thus, trade volume is moving away from the NYSE even as the performance gap between the NYSE and non-NYSE trade is widening. Despite its superior price performance, volume on the Cincinnati exchange actually decreases between 1988 and 1989. Meanwhile, the fastest volume increase is at the NASD, the market with the least favorable execution costs.[20]

VI. Price Improvement Tests

Panel A of Table VII reports the results of price improvement tests, based on the percentage of trades inside the best ITS spread for each exchange. The NYSE trades in the match sample are executed inside the spread 39 percent of the time in 1988 and 37 percent in 1989. By comparison, the Cincinnati trades are inside the spread 48 and 49 percent of the time in 1988 and 1989, respectively. NASD trades are inside the spread only 30 and 27 percent of the time in 1988 and 1989. As expected, the price improvement observed in the earlier tests is related to the frequency of inside-the-spread trading at each location.

Since much of the order flow targeted for purchase is executed when the spread is one eighth, these trades provide little opportunity for price improve-

[20] The flow of trading volume is not a comment about investor rationality. Investors may be fully rational, but not informed with respect to this aspect of their execution costs. Alternatively, investors may be attracted to the regional markets by factors other than price execution.

Table VI

Estimated Additional Cost of Off-Board Executions

This table presents the estimated total additional execution costs incurred by investors who had their trades executed at regional exchanges and OTC markets during 1988 and 1989. The Panel A results are based on the excess liquidity premium paid relative to adjacent NYSE trades of similar size. The Panel B results are based on the average additional cost of buying-selling relative to adjacent NYSE buys-sells of similar size. NYSE trades are deemed adjacent if they occur within two minutes of the regional or OTC trade.

Panel A: Based on the Excess Liquidity Premium Test

	1988			1989		
	Average Excess Liquidity Premium (¢/Share) from Table IV	Volume During Year (Millions of Shares)	Additional Cost (Millions of Dollars)	Average Excess Liquidity Premium (¢/Share) from Table IV	Volume During Year (Millions of Shares)	Additional Cost (Millions of Dollars)
Boston	1.20	593.1	7.09	0.91	799.6	7.08
Cincinnati	0.43	251.9	1.08	1.62	228.6	3.71
Midwest	−0.20	2632.8	−5.20	−0.01	2784.4	−0.22
Pacific	0.01	1329.8	0.08	0.48	1536.9	7.43
Philadelphia	0.35	631.7	2.22	1.28	910.2	11.69
NASD	1.10	1034.1	11.43	1.29	1594.5	20.64
Instinet	−4.02	67.3	−2.71	0.81	67.5	0.55
All	0.20	6540.7	12.97	0.59	7901.7	46.96

Table VI—*Continued*

Panel B: Based on the Directional Price Test						
	1988			1989		
	Average Additional Cost to Buy and Sell (¢/Share)*	Volume During Year (Millions of Shares)	Additional Cost (Millions of Dollars)	Average Additional Cost to Buy and Sell (¢/Share)*	Volume During Year (Millions of Shares)	Additional Cost (Millions of Dollars)
Boston	0.93	593.1	5.52	0.84	779.6	6.54
Cincinnati	0.44	251.9	1.11	0.21	228.6	0.47
Midwest	−0.09	2632.8	−2.34	0.07	2784.4	1.93
Pacific	−0.02	1329.8	−0.23	0.27	1536.9	4.12
Philadelphia	0.61	631.7	3.82	0.92	910.2	8.41
NASD	1.35	1034.1	13.96	1.23	1594.5	19.65
Instinet	0.13	67.3	0.09	1.18	67.5	0.80
All	0.28	6540.7	17.99	0.46	7901.7	36.32

* Computed as the average additional cost to buy and sell.

Table VII

Relative Performance by Frequency of Inside-the-Spread Trading

Panel A presents the frequency of inside-the-spread trades as a percentage of all trades in the matched sample for that exchange. Panel B table values represent the percentage of trades at a given spread size executed inside the spread. As such, Panel B table values capture the probability that an execution will be inside the spread, conditional on location and spread size. In both panels, the spread is computed using the best available ITS quote at the time of the trade execution.

	Spread Size (1988)				Spread Size (1989)			
	1/4	3/8	1/2 or More	All	1/4	3/8	1/2 or More	All
	Panel A: Percentage of Trades Inside the Spread By Exchange and Size of Spread							
NYSE	19.2	10.9	9.4	39.5	20.0	9.6	7.7	37.3
Boston	15.4	10.6	11.0	37.0	16.1	9.7	8.5	34.3
Cincinnati	21.1	14.4	12.1	47.6	22.8	14.1	11.8	48.8
Midwest	16.9	10.8	9.9	37.6	17.7	10.2	8.4	36.3
Pacific	16.7	10.6	9.8	37.1	17.4	9.9	8.5	35.8
Philadelphia	13.3	9.9	8.6	31.9	13.5	8.4	7.3	29.2
NASD	13.3	9.2	7.7	30.1	13.4	7.3	6.6	27.4
Instinet	17.9	10.9	11.9	40.7	16.1	12.5	9.5	39.1
	Panel B: Inside-the-Spread Trades as a Percentage of Total Trades at Each Spread Size							
NYSE	61.6	66.9	62.7		64.7	68.6	65.4	
Boston	44.2	64.3	62.5		45.6	66.9	62.0	
Cincinnati	61.6	59.9	49.5		62.4	61.9	48.9	
Midwest	47.2	66.2	65.7		49.3	68.4	66.4	
Pacific	47.8	65.0	63.1		48.8	67.8	64.8	
Philadelphia	37.7	59.3	60.7		38.1	59.0	59.2	
NASD	38.0	60.8	64.8		39.0	59.4	62.2	
Instinet	44.9	65.9	77.3		47.5	67.2	69.3	

ment (Stern (1989)). For example, if relatively more NASD trades are executed when the NYSE spread is only one eighth, the results in Panel A might simply reflect fewer opportunities for price improvement at the NASD. To address this possibility, Panel B reports the inside-the-spread trades as a percentage of the total number of trades at each spread size. Thus, Panel B measures the likelihood of price improvement at each location, conditional on the spread size.

Two results are striking. First, when the best ITS quote is greater than one eighth, a substantial proportion of the trading takes place inside the spread. At a one-quarter spread, a full 62 percent (or 38 percent) of the NYSE (or NASD) trades execute between the best bid and ask prices. This finding shows the best ITS quote is not capturing all the liquidity available in the market. Second, the relative ranking across exchanges is the same as in the first two tests. The market centers that are most likely to improve on quoted prices (NYSE, Cincinnati, Midwest, and Pacific) are the locations with the best price execution. The NASD is least likely to improve prices relative to the ITS quote, even when price improvement is possible.

VII. Supplemental Analyses

A. Size of the Match Window

All the reported results are for trades matched on the basis of a two-minute window. A wider window provides larger sample sizes, but less assurance that market conditions are similar for NYSE and off-Board trades. To examine the sensitivity of the results to this parameter, I recompute the excess liquidity premiums using one-, two-, five-, and ten-minute windows. Since the results do not vary significantly by window size, they are not reported separately. In general, as the window size decreases, the NYSE performance tends to improve against the performance of other markets. However, intermarket rankings are largely unchanged.

B. Timing Issues

A potential concern with the research design is the mismatching of trades due to intermarket differences in the timeliness of reporting. If matching NYSE trades are systematically recorded ahead of the corresponding regional trades (for example, if NYSE traders tend to respond faster on both upward and downward price moves), the earlier results could be biased in favor of the NYSE. The two-minute match window, while mitigating the problem, may not fully alleviate it. To examine this possibility, Figure 2 shows the distribution of adjacent NYSE trades for the ten minutes around a regional trade. The slight increase in the NYSE volume in the same minute as the regional trades suggests the arrival time of orders across the exchanges is correlated. However, matched NYSE trades do not appear to be executed earlier than their regional counterparts. In fact, the regional trades have a slight speed advantage over their NYSE control sample.

Another timing issue is the problem of identifying the appropriate quote for each trade. Lee and Ready (1991) show trades that trigger quote revisions may be recorded on the ISSM tape after the quote revision. This is because the NYSE quote revision is typically entered electronically, while many trades are still recorded using the slower card readers. Since the new quote tends to bracket the trade that triggered it, the triggering trade will have the appearance of price improvement. If NYSE trades are more likely to trigger quote revisions, the reported results might reflect this bias. The five-second quote delay can mitigate this problem, but not resolve it.

Ideally, this problem should be addressed by identifying the prevailing quote at the time each order is submitted. Since this is impossible with the ISSM data, an alternative test is performed. Specifically, the liquidity pre-

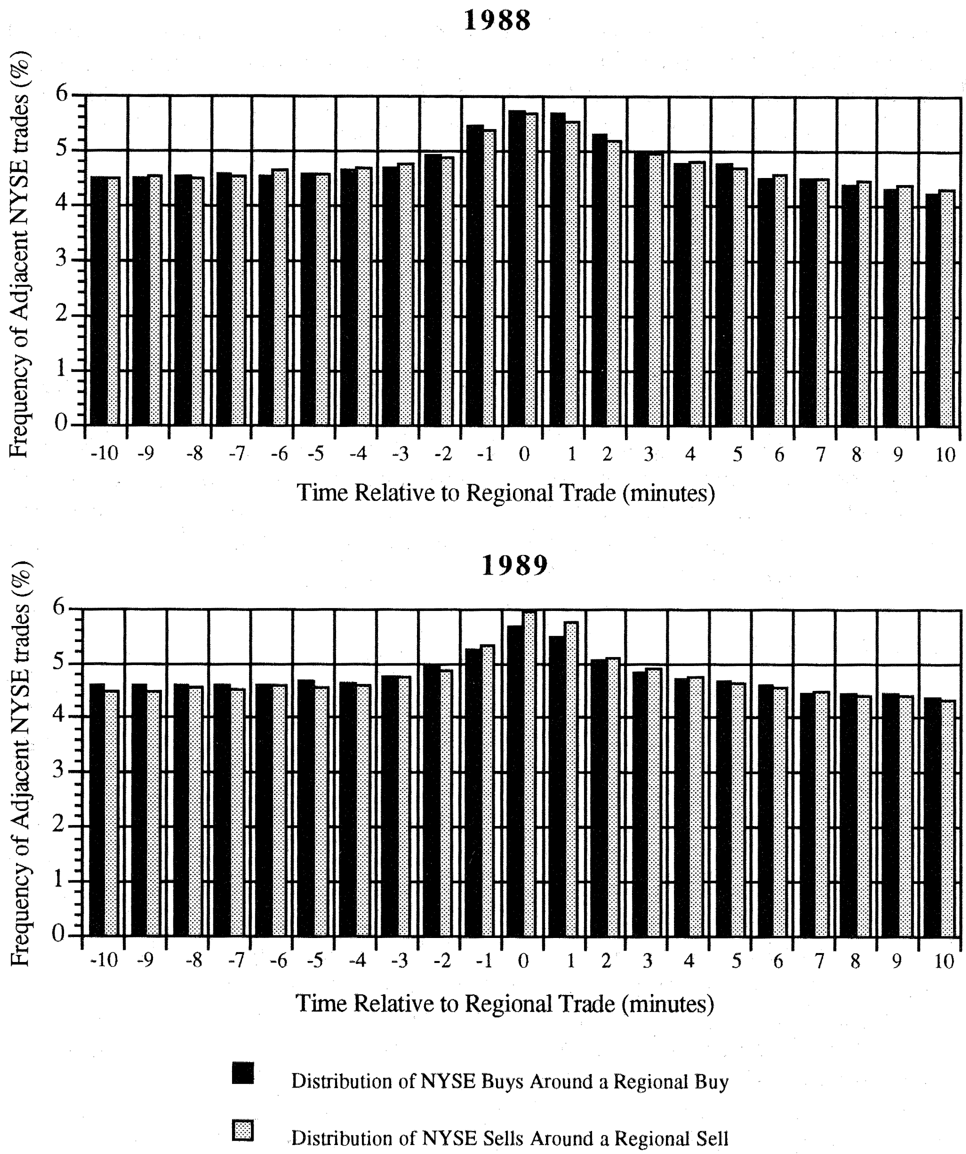

Figure 2. The distribution of adjacent NYSE trades of similar size and direction around a regional trade.

mium test is repeated twice, using the prevailing best ITS quote at least two minutes and at least five minutes before each trade. Some of these quotes are superseded by newer quotes, so this approach will introduce noise into the calculation. However, it provides less risk of a bias toward a particular exchange. The results (not reported) remain similar, with no change in the intermarket rankings.

C. Regional Trades Due to "Limit Book Protection"

The regional exchanges and the NYSE handle public limit orders differently in one important respect. On the NYSE, public limit orders have the opportunity to be exposed to a large volume of incoming market orders. To ensure that public limit orders at the regionals receive comparable execution, regional market makers typically guarantee execution of these orders when the limit order price is penetrated by a transaction in another market center. To implement this limit book protection, the regional market maker initiates a "shadow" trade at his own exchange for the same price as the triggering trade. This trade triggers the regional limit order, thus ensuring it receives as good an execution as the limit orders on other exchanges.

Since these shadow trades are not initiated by public market orders, the argument could be made that they should be excluded from the analysis. This exclusion is not possible with ISSM data, nor is the extent of this activity easily quantified. However, the likely effect of including these trades is to understate reported intermarket differences. Since regional shadow trades are typically triggered by a NYSE trade at the same price, paired observations caused by limit book protection should show identical execution costs. Therefore, the inclusion of these observations in the sample reduces the average price difference between NYSE and non-NYSE executions.

D. Accounting for Cross-Sectional Firm Characteristics

We should note that these tests do not control for cross-sectional differences across the sample firms. The unit of analysis is individual trades, so that securities with greater volume receive greater proportional weight. This approach is reasonable given the research focus on price execution per trade. However, since the extent of off-Board trading varies widely across securities, the results may be due to a small proportion of the sample firms. To examine this possibility, the liquidity premium tests are recomputed so that the excess cost per trade is first averaged for each firm, then aggregated across firms. Statistical significance is measured by the proportion of firms in which the NYSE (or the regional exchange) demonstrates superior price performance. Once again, since the findings for these tests confirm the earlier results, they are not separately reported.

E. Other Potential Matching Problems

Finally, since no two orders are identical in all respects, other matching problems could remain between NYSE and non-NYSE trades. For example, larger market orders are sometimes broken up in execution as they march through the limit book. Since the NYSE tends to receive large orders, a large trade sent to the NYSE may appear as a series of smaller trades, all poorly executed relative to the prevailing quote. In this case, the price performance of the NYSE would appear worse. In other situations, regional price performance could be slighted. Short of a controlled laboratory experiment, such differences are difficult to eliminate. Therefore, this study should be regarded as a first step in the continuing process of understanding market integration and intermarket price performance.

VIII. Summary

This paper presents three simple ways to measure the execution and price performance of different market centers. The results show that for NYSE-listed securities the price obtained on similar adjacent trades can differ by location of execution. In particular, the results for 1988 and 1989 suggest that Cincinnati, Midwest, and NYSE executions are generally more favorable to the initiator of the trade than executions on the other exchanges. The average price difference between the NYSE and matched off-Board trades is 0.7 to 1 cent per share. These results are statistically significant at any reasonable level of confidence. In total dollars, the aggregate excess cost for off-Board trades is estimated at $13 to $18 million for 1988 and $36 to $47 million for 1989.

Several implications derive from these results. First, the existing intermarket structure for NYSE securities is much less integrated than many had previously assumed. These findings suggest that order routing may warrant consideration by investors and regulators. Brokers currently rely on the ITS quote to fulfill "best execution" obligations. This strategy may be inappropriate given the effects of market fragmentation documented here. Many investors currently delegate the responsibility of choosing the best market center to their broker. These results suggest individual investors may wish to consider other arrangements, discussed below, to ensure the benefits of order inducements flow directly to them.

These results also suggest that full integration may be difficult to attain under the current ITS system. Under ITS, liquidity providers at the primary exchange may not be motivated to reveal their presence to off-Board traders. Thus, the current system encourages gamesmanship between exchanges, as each probes the other for "hidden liquidity" not revealed in the intermarket quote. The price improvement algorithm recently implemented by a prominent NASD member firm, Madoff Investments, is a prime example of this type of intermarket price exploration (see Appendix A). A long-term solution to the market integration problem may involve more fundamental changes to the intermarket quote system. One possibility is a consolidated electronic limit book with strict price and time priority. This appears to be a promising area for future research.

The order flow debate involves factors not considered in this study. However, these results do raise several interesting issues pertinent to the debate. First the magnitude of the price differences is comparable to order flow cash inducements. In light of these price differences, the broker's primary responsibility could potentially be compromised by order flow inducements. Second, the performance gap between NYSE and off-Board executions is most evident in the small trades. This finding is disturbing, since small trades are the primary target of order flow inducements and small traders may be least aware of the dealer-broker arrangements. Finally, some (limited) evidence suggests that trading volume is not flowing in the direction of better price execution. Prices obtained on the NASD are typically least favorable, yet the trading volume on this exchange has increased most dramatically in recent years. In contrast, the better performing centers, such as Cincinnati and the NYSE, actually experienced a decline in market share between 1988 and 1989. Investors may be moving to the NASD for nonpecuniary reasons, but the volume shift may also reflect broker response to cash inducements. The finding is at least suggestive that order flow patterns may be responsive to cash inducements, rather than price execution.[21]

While investors appear to pay more for off-Board executions, brokers who receive order flow inducements might indirectly pass on these savings through lower commission fees. Given competition among brokers, investors may be partially compensated for the additional costs through lower broker commissions. However, investors whose orders are diverted are unlikely to be fully compensated for two reasons. First, unless enough traders exercise their right to direct orders, brokers are not compelled to pass on all the inducement benefits to investors. Second, even if all the inducement benefits are reflected in the commission structure, these savings are spread out across all investors, so investors who have their orders diverted still subsidize those who do not.

More importantly, investors should recognize that dealer and broker services are fundamentally separable. Traders that engage a broker do not legally commit themselves to a particular dealer. Since investors can shop for both the lowest broker commission and the lowest dealer (execution) cost, these costs are not fungible. Currently, many investors delegate the choice of a dealer to their broker, presumably because the choice appears inconsequential. The evidence here suggests some investors—e.g., small traders—may be better off not delegating this decision. An investor could either direct an order to a particular location or "sell" this right by choosing a low-priced broker who does not allow location choices. In either case, the benefit of order flow inducements passes directly to the ultimate consumer. These arrangements reflect the fact that order flows belong to investors, not brokers.

Finally, the existence of intermarket price differences suggests brokers should be called to higher level of accountability for their trade execution performance. A simple first step is to require brokers to report the location where the order is executed. Another step is to make available periodic measures of comparative execution costs by location, and perhaps even by brokerage house. Given information on the comparative costs of alternative executions, informed investors can decide for themselves whether to take further action. As the consuming public focuses on these measures of perfor-

[21] Two years of data cannot provide strong conclusions. Future studies should be able to provide more evidence on this issue.

mance, the various parties competing for their business should modify their actions to provide the most favorable execution possible.

Appendix A: Electronic Systems at Regional Exchanges and the NASD

Most regional and OTC markets have their own price improvement procedures. These typically involve guaranteeing the best ITS price and further exposing an incoming order to either the floor of the local exchange or the ITS system for a period of time (i.e., "stopping" a trade). While incoming orders may be manually "stopped" at any time, electronic systems at many exchanges automatically "stop" certain types of orders. These automated systems are briefly summarized below.

Boston, Pacific, and Midwest

These three exchanges have similar systems. The BEACON (Boston), Scorex (Pacific), and MAX (Midwest) systems all display orders of up to 1,099 shares for 15 seconds to the local floor for possible improvement. If a better price is not obtained in 15 seconds, the order is executed against the best intermarket quote at the time of receipt.

In addition, during 1990, the Midwest exchange introduced a new price improvement algorithm that offers a one-eighth improvement on all trades if its execution at the intermarket quote generates a double down or double up tick (i.e., two successive negative or positive price changes).

Cincinnati

Cincinnati has no exchange floor and operates solely on a quote display system. The NSTS system at Cincinnati automatically executes orders of up to 2,099 shares at the best intermarket quote.

Philadelphia

The PACE system of Philadelphia automatically executes at the best intermarket quote for up to 599 shares.

NASD

Most NASD market makers during this study period automatically executed orders at the best intermarket quote. However, during the latter part of 1990, Madoff Investment (a prominent member of NASD), instituted an automated execution system dubbed MISSION. For orders of 300 shares or greater, in stocks trading at one-fourth or more spread, MISSION adjusts the Madoff ITS bid or ask to reflect the customer's order for one minute. If this quote is taken within the one minute, the entire order is cleared at the improved price.

Appendix B: Inferring Trade Direction

The direction of individual trades is inferred by the following algorithm developed in Lee and Ready (1991). The only modification is that Lee and Ready use only NYSE quotes, while this study uses all BBO-eligible quotes on the ISSM tape (a quote is BBO-eligible if it qualifies for the NASD Best-Bid-Or-Offer calculation):

1. Current Quote Match—If the trade price is at the bid or ask, and the current quote was not revised within the last 5 seconds, then the direction of the trade is determined by the current quote (i.e. a buy if it is at the ask and a sell if it is at the bid).
2. Delayed Quote Match—If the current quote is less than 5-seconds old, it is ignored and the trade price is compared to the bid and ask prices of the previous quote.
3. Outside the Spread—If the trade price, when compared to the quote in either 1 or 2, is greater than the ask (or less than the bid), then the transaction is deemed a buy (or sell).
4. Tick Test—If the trade is at the midpoint of the spread, or if a BBO-eligible quote is not available, the tick test is used to determine trade direction. A BBO-eligible quote is deemed to be unavailable if the last NYSE-quote issued has a nontradable condition code. Using the tick test, if the last price change was positive (or negative), then the current trade is deemed a buy (or sell). All out-of-sequence trades are ignored in updating price changes.
5. Proximity to Bid-Ask—If a trade is between the spread but not at the midpoint, then the trade is classified according to its proximity to the bid or ask price. Trades at prices above the midpoint are classified as buys and trades at prices below the midpoint are classified as sells.
6. Indeterminable—This classification is assigned to a trade when none of the above conditions apply. Specifically, it applies to the first trade of the year for each firm and any trade which is reported out of sequence.

REFERENCES

Blume, Marshall E., and Michael A. Goldstein, 1991, Differences in execution prices among the NYSE, the regionals and the NASD, Working paper, University of Pennsylvania.

Coffee, John C., Jr., 1991, Brokers and bribery, *New York Law Journal*, September 27, 5.

Crawford, William B., Jr., 1991, Brokerage fees stir debate on exchange floors, *Chicago Tribune*, August 26.

Easley, David, and Maureen O'Hara, 1987, Price, trade size, and information in securities markets, *Journal of Financial Economics* 19, 69–90.

Glosten, Lawrence R., 1991, Asymmetric information, the third market and investor welfare, Working paper, Columbia University.

——— and Paul Milgrom, 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 14, 71–100.

Harris, Lawrence E., 1990, *Liquidity, Trading Rules, and Electronic Trading Systems,* New York University Monograph Series in Finance and Economics, No. 1990–4.

———, 1992, Consolidation, fragmentation, segmentation and regulation, Working paper, University of Southern California.

——— and Joel Hasbrouck, 1992, Market vs. limit orders: The SuperDOT evidence on order submission strategy, Working paper, University of Southern California and New York University.

Hasbrouck, Joel, 1988, Trades, quotes, inventories and information, *Journal of Financial Economics* 22, 229–252.

——— and George Sofianos, 1992, The trades of market-makers: An analysis of NYSE specialists, Working paper, New York University and New York Stock Exchange.

Lee, Charles M. C., Belinda Mucklow, and Mark J. Ready, 1993, Spreads, depths and the impact of earnings information: An intraday analysis, *Review of Financial Studies* 6, Forthcoming.

Lee, Charles M. C., and Mark J. Ready, 1991, Inferring trade direction from intraday data, *Journal of Finance* 46, 733–746.

Mann, Steven V., and Robert W. Seijas, 1991, Bid-ask spreads, NYSE specialists, and NASD dealers, *Journal of Portfolio Management*, 17, 54–58.

Mayer, Marcia Kramer, and Bradford Leigh, 1991, Does off-board trading compromise execution quality? The evidence from high and down low, Working paper, American Stock Exchange Research Department.

National Association of Securities Dealers, 1991, *Inducement For Order Flow, A Report of the Order Flow Committee to the Board of Governors* (National Association of Securities Dealers, Inc., Washington, D.C.).

Perold, Andre F., 1988, The implementation shortfall: Paper versus reality, *Journal of Portfolio Management* 14, 4–9.

Petersen, Mitchell, and Steve Umlauf, 1991, An empirical examination of intraday quote revisions on the New York Stock Exchange, Working paper, University of Chicago.

Reinganum, Marc R., 1990, Market microstructure and asset pricing, *Journal of Financial Economics* 28, 127–147.

Securities Week, 1989a, NASD continues studying payment for order flow study, asks for more time, May 8.

———, 1989b, Industry ponders question of exchanges paying for order flow, March 6.

Steptoe, Sonja, 1989, Debate over order flow payments grows hotter as competition increases, *Wall Street Journal*, February 21.

Stern, Richard, 1989, Living off the spread, *Forbes*, July 10, 66–67.

Wood, Robert A., Thomas H. McInish, and J. Keith Ord, 1985, An investigation of transactions data for NYSE stocks, *Journal of Finance* 40, 723–739.