Journal of Economic Perspectives—Volume 20, Number 1—Winter 2006—Pages 153–174

Electronic Trading in Stock Markets

Hans R. Stoll

In 1792, 24 brokers met under a buttonwood tree to found what became the New York Stock Exchange (NYSE). Thereafter, they carried out transactions in a coffeehouse located at Wall and Water streets in lower Manhattan. Transactions were consummated via face-to-face negotiation. From its beginning, the NYSE was a member organization—a club—in which trading could be conducted only through members at uniform fees established by the members. Each member was required to purchase a membership, or seat, in order to do business on the exchange. Today there are 1,366 seats. Trading was conducted by hand-carrying orders to the appointed trading post where trading by members would take place.

Modern trading technology clashes with the traditional organization of a stock exchange. The facility—the coffeehouse—has changed. The modern facility is no longer a place. Rather, it is a computer system over which transactions are entered, routed, executed and cleared electronically with little or no human intervention. The role of the brokers and dealers has changed, and the need for a physical meeting place has become of little importance. Competition from fully electronic markets—electronic communications networks (ECNs)—and from regional exchanges—Boston, Chicago, Philadelphia, Pacific, Cincinnati—as well as regulatory pressures, are forcing the NYSE to adapt. As it does, the role of members as brokers and dealers will diminish and their role as owners and operators of the technology will increase. The clearest reflection of change is the NYSE's announcement, on April 20, 2005, that it intended to become a publicly owned company and to merge with Archipelago, one of the most successful ECNs. The Nasdaq Stock

■ *Hans R. Stoll is the Anne Marie and Thomas B. Walker Professor of Finance and Director of the Financial Markets Research Center, Owen Graduate School of Management, Vanderbilt University, Nashville, Tennessee. His e-mail address is* ⟨*hans.stoll@owen.vanderbilt.edu*⟩.

Market, facing even greater competitive pressures than the NYSE, has announced a merger with Instinet, the other major ECN.[1] As a consequence of these pending mergers, one can expect increased electronic trading in all markets.

In this article, I examine how electronic trading has altered stock markets. I begin with an overview of how the stock trading process works and then address a number of questions. How have the jobs of traditional stock market dealers on the NYSE and on Nasdaq been affected by electronic trading? How do electronic communications networks differ from traditional markets? How has electronic trading affected bid-ask spreads and commission costs? What subtle issues arise in electronic trading when dealer and customer interests diverge? Will computer programs replace human judgment? What is the effect of electronic trading on the number and types of securities markets? What is the role of regulation in electronic markets?

The Stock Trading Process

A stock market is the simplest type of financial intermediary. Unlike a bank, which transforms deposits into loans, a stock market only transfers ownership from one investor to another. This task would be easy if buyers and sellers arrived at the same time and wanted to trade the same amounts at the same prices, but they do not. Consequently, markets must have the facility to search for the other side of trades and to provide liquidity should the other side not be available. Computer technology affects the types of markets that are most efficient in carrying out these tasks.

The trading process begins when investors enter orders with their brokers to buy or sell shares of stock. The basic order types are a market order, which requires the broker to trade immediately at the best possible price, or a limit order, which specifies a maximum price if buying or a minimum price if selling. The broker next routes the order to a particular market. Traditionally, orders in NYSE-listed stocks were routed to the NYSE while orders in Nasdaq stocks were routed to the Nasdaq Stock Market. Today the connection between listing and trading of stocks is weakening. Stocks may be dually listed on more than one market, but the primary listing venue remains either the NYSE or Nasdaq and, to a lesser extent, the American Stock Exchange (AMSE). Table 1 presents some basic information about the exchanges on which stocks are traded.

Stocks may be traded in markets other than where they are listed under the unlisted trading privilege (UTP) rule of the Securities and Exchange Commission. For example, about 20 percent of the volume in NYSE stocks is traded in other markets (as shown in Table 1). The Nasdaq market trades only about 51 percent of

[1] "Nasdaq" originally stood for "National Association of Securities Dealers Automated Quotation." Nasdaq was founded in 1971 as a dealer quotation system. The electronic communications network is actually INET, a subsidiary of Instinet. We shall refer to it as Instinet.

Table 1
Where are Stocks Listed and Traded in 2004?

	NYSE	*Nasdaq*	*AMSE*	*Regional*	*ECN*
Number of stocks listed (primary listing)	2,780	3,200	650	[a]	0
Dollar market value of listed stocks ($ billion)	17,800	3,700	53	[a]	0
Percentage share volume in NYSE listed stocks[b]	79%	14%	0%	4%	3%
Percentage share volume in Nasdaq listed stocks[b]	0%	51%	0%	7%	42%

Source: NYSE and Nasdaq Stock Market websites, Securities and Exchange Commission Annual Report.
Notes: [a]Regional exchanges primary listings are minor and contribute little to volume. Almost all their volume is in stocks with primary listing on the NYSE or on Nasdaq.
[b]Trading volume is reported by the NYSE, Nasdaq and the regional exchanges under a consolidated tape reporting system. The regional exchanges are Pacific, Chicago, Philadelphia, Boston and Cincinnati. ECN volume is not reported separately, but it is known that Archipelago reports via the Pacific Exchange (which it owns), and Instinet reports via the Cincinnati Exchange. I assume that ECN volume is the sum of Pacific and Cincinnati volume and that regional volume is the sum of volume on Chicago, Philadelphia and Boston.

volume in stocks it lists. About 42 percent of volume in Nasdaq stocks is traded on the two largest ECNs, Archipelago and Instinet. Thus, brokers have options beyond the listing market where orders may be routed. While the mergers of Archipelago and the NYSE and of Insinet and Nasdaq will reduce those options, it is expected that competition between the NYSE and Nasdaq to trade each other's stocks will intensify. After an order has been routed to a particular market, it is executed and a confirmation is sent to the customer. If the entire process is fully automated, then the steps from order entry to order routing, execution and confirmation can be completed in seconds, perhaps microseconds. The final step in the trading process is to transfer share ownership and make payment. In U.S. stock markets, this last step is completed within three days via computer book entry at the clearing agency, the Depository Trust and Clearing Corporation.

The most critical step in the trading process is the execution of the order, for at this point the price is determined. The exact process by which the transaction is executed and the price determined varies across markets and across the time of day. The opening on many markets, including the NYSE, operates according to a call auction procedure where the orders arriving before the open are batched and executed at a single opening price. A computer algorithm calculates an opening price that maximizes the volume of trading.

Once the market is open, trading takes place continuously as arriving orders are traded in sequence. In a continuous *dealer* market, incoming market orders trade at bid and ask quotes placed by dealers. For example, if the bid price at which dealers are willing to buy is $30.00 and the ask price at which dealers are willing to sell is $30.10, investors wishing to sell would sell at $30.00, and investors wishing to

buy would buy at $30.10. Dealers trade for their own accounts as principals, in contrast to brokers who act as agents, and dealers earn income from the difference between the bid price at which they buy and the ask (or offer) price at which they sell. Dealers are sometimes referred to by the term "market-makers." Bond markets, currency markets and the Nasdaq market are primarily dealer markets. In the Nasdaq market, at least two dealers (or "market-makers") make markets in each stock, with an average of about six dealers per stock. The bid-ask spread is a measure of the cost of immediacy, a term coined by Demsetz (1968). Dealers supply immediacy and earn the spread, while demanders of immediacy pay the spread. The spread compensates dealers for the risks they assume in buying (or selling) a stock that someone else is anxious to sell (or buy), as well as for the costs of processing the trade. The spread may also reflect monopoly rents if dealers have market power.

In a continuous *auction* market, incoming market orders trade at prices placed by investor limit orders that have been entered into a "book." The limit prices in the book constitute the bid and ask prices in an auction market. The advantage of a continuous auction market is said to be that investors trade directly with other investors. In fact, a pure continuous auction market is unworkable for many securities because an insufficient number of limit orders are placed to maintain liquidity. Exchange markets, like the NYSE and the AMSE are mixed dealer/auction markets in which limit orders are supplemented by the liquidity supplied by the specialist. Each stock on the NYSE is assigned to a specialist, who is responsible for "maintaining a fair and orderly market" in the stock. The specialist acts in a dual capacity: as a broker representing the limit orders left with him and entered in the display book and as a dealer trading for his own account. The specialist also orchestrates the opening auction and oversees trading by brokers on the floor.

Table 2 gives an example of a book of limit orders sorted by price. Before computer technology, limit orders would be entered manually into a book. Today, orders are entered in the computer and displayed on a screen. On the NYSE, execution of an order usually requires action by the specialist to specify which incoming order trades with which limit order on the book (or with which broker on the floor). On electronic communications networks, the entire trading process is automatic. Incoming limit orders are automatically entered, displayed and executed in the computer against incoming market orders. For example, the book for ABC in Table 2 indicates that the most any limit order will pay is $20.87 and that up to 1,000 shares would be purchased at that price. A market order to sell up to 1,000 shares would trade with this limit order at $20.87. A market order to buy would trade with the lowest ask price, which is $20.93. In a very active stock, orders arrive very rapidly and small price changes occur so frequently that the best price is fleeting—the computer screen flickers rapidly from one price to the next. In less active stocks, updates can be infrequent, and dealers or other traders may be necessary to provide liquidity.

Table 2
Limit Orders in ABC Company Stock

Bid		*Ask*	
Price	*Quantity*	*Price*	*Quantity*
20.87	1000	20.93	1000
20.86	1500	20.94	100
20.83	1500	20.99	300
20.74	200	21.00	2000
20.72	100		

The NYSE Specialist

The specialist is central to the operation of the NYSE. As a dealer, the specialist stabilizes stock prices by buying when the public is selling and selling when the public is buying. As a broker, the specialist represents orders left on the book of limit orders and executes them when they represent the best price. That central role has not changed much despite the rise of electronic trading. For example, in 1975, specialists acting as dealers participated in 25.4 percent of the volume.[2] In 2003, the participation rate was 25 percent. However, the way in which trading takes place has changed. Most of the orders on the NYSE arrive electronically over SuperDot, a system that routes orders directly to the specialist's display book. In 2002, over 98 percent of all orders and 70 percent of trade volume were routed over the SuperDot system. The balance of orders is in the form of large transactions that are negotiated "upstairs" over the telephone or traded by brokers on the floor.

Technology has changed the nature of the specialist in other ways, too. In 1975, 381 individual specialists owned seats and operated 67 specialist firms organized primarily as partnerships. Today, the number of individuals acting as specialists remains about the same, but they are organized into only seven specialist firms, structured as corporations. The consolidation of dealer firms on the NYSE (and also in other markets) reflects the economies of scale in the dealer business.

The seats not owned by specialists are held by floor brokers or by brokerage firms sending orders to the NYSE. Floor brokers trade on the floor as agents for institutional investors with more difficult trades than can be readily handled on SuperDot. However, as the capability of electronic markets continues to improve, traditional floor brokers are threatened. Retail brokerage firms that wish to send orders directly to the NYSE, rather than going through a member firm, must themselves be members. The desirability of being an NYSE member, either as a

[2] Specialists accounted for 12.7 percent of shares sold and 12.7 percent of shares purchased. Since specialists would not be on both sides of the same trade, they participated on one side or the other in 25.4 percent of the trading volume in 1975. Data are from the NYSE Fact Book. Stoll (1985) estimates that the specialist acted as broker for one side or the other in 52 percent of the volume in 1975. The role of the floor broker is analyzed by Sofianos and Werner (2000).

specialist, floor broker or retail broker, is reflected in seat prices. As alternative electronic markets grow, the need for NYSE access declines, and NYSE seat prices would be expected to fall. NYSE seat prices did fall to $975,000 in January 2005, but have recovered dramatically to $3 million by August 3, 2005 (Ceron, 2005), an increase that reflects the anticipated benefits of going public and merging with Archipelago.

Specialists are frequently criticized for the conflict inherent in their dual roles as brokers for the orders left with them to be entered in the book and as dealers trading for their own accounts. Their knowledge of the book and their central role in the trading process give them a trading advantage that might be used to their own benefit at the expense of orders for which they act as brokers. Securities and Exchange Commission and NYSE regulations prohibit actions by the specialist that disadvantage investors, and they impose both negative and affirmative obligations on the specialist. The negative obligation limits specialist trading to that necessary for the maintenance of a fair and orderly market. The affirmative obligation requires specialists to trade for their own accounts to maintain price continuity and stabilize prices. Whether these regulatory obligations are desirable or whether eliminating the privileged position of the specialist would be a better approach remains a matter of debate.[3]

The Securities and Exchange Commission's most recent investigation of specialists found that certain individual specialists traded ahead of customer orders in violation of the negative obligation not to trade "unless reasonably necessary." The firms paid a total of $242 million to settle the charges against them (Securities and Exchange Commission, 2004). On April 12, 2005, 15 individual specialists who worked for the firms were indicted, and the NYSE paid a fine of $20 million for failure to regulate the specialists properly (Anderson, 2005). One type of violation cited by the Securities and Exchange Commission was trading ahead of customer orders. For example, if the specialist observed that a disproportionate number of buy orders were arriving, he could delay execution of these orders, buy for his own account at a lower price and then sell to the incoming orders at a higher price. A second type of violation was "interpositioning." The Securities and Exchange Commission concluded that specialists would unnecessarily interposition themselves between arriving buy and sell market orders from customers rather than letting customer orders trade against each other. Suppose the limit order book shows a bid of $20.87 and an ask of $20.93. As market orders arrived they would normally trade with the book (at $20.87 if the market order were a sale and at $20.93 if the market order were a buy). Alternatively, if market buy and sell orders arrived simultaneously, they could be paired against each other at a price between $20.87 and $20.93. Instead, the Securities and Exchange Commission found that the specialist would buy from the incoming sell order at $20.88 and sell to the

[3] Stoll (1998) argues for eliminating the special privileges and obligations of the specialist.

incoming buy order at $20.92, thereby earning a spread, an unnecessary intervention because the buy and sell orders could have been paired off inside the spread.

The source of these problems is that the execution of trades is not automatic. The specialist has discretion about when and how orders are executed and when he participates. Market orders can, for example, be allowed to accumulate rather than be executed automatically against the prevailing quote. As they accumulate, the specialist has the discretion to decide whether to 1) trade ahead of the orders; 2) let the orders trade against the book; 3) trade with the orders for his own account; or 4) pair them off within the spread. This discretion benefits the specialist. No such discretion would exist in a fully automated market where executions occurred automatically and immediately at the best price.

The Nasdaq Market

The Nasdaq Stock Market owes its very existence to computer and communications technology. Before 1971, less prominent stocks that were not listed on the New York Stock Exchange traded in an "over-the-counter" network of dealers, or market-makers, who distributed hard copy "pink sheets" that quoted indicative but not binding bid and ask prices. In 1971, Nasdaq began electronic dissemination of this information, and trades would be consummated by telephone negotiation based on the disseminated quotes. Over time, quotes became "firm" and could be directly accessed, and features of exchange markets, like last sale reporting, were introduced. By 1990, the Nasdaq Stock Market was firmly established. It listed more stocks than the NYSE and retained companies like Intel, Microsoft and MCI that in past years would have listed on the NYSE. Trading in Nasdaq stocks took place electronically over computer and telephone lines connecting the offices of about 500 market-maker firms and numerous order entry firms. Each market-maker firm would disseminate quotes in the stock in which it was a market-maker. Unlike the NYSE, where there was only one market-maker (the specialist) per stock, Nasdaq stocks were required to have at least two market-makers. Active stocks attracted more than 40 market-makers.

In view of the number of competing market-makers and the apparent high-tech and transparent nature of the Nasdaq market, it might seem that transactions costs would be reduced with better prices for investors. However, all was not well. Several academic studies and investigations by the Securities and Exchange Commission and the Justice Department revealed dealer behavior that artificially raised bid-ask spreads above competitive levels (see in particular Christie and Schultz, 1994, 1995; Christie, Harris and Schultz, 1994; Huang and Stoll, 1996). The source of the problem was that in a pure dealer market, which Nasdaq resembled at that time, customers must trade at dealer quotes. Customers could not trade with another customer's limit order that was at a better price. Consequently, the bid-ask spread was determined by dealers without the possibility of competition from customers.

For several reasons, dealers did not compete effectively among themselves to narrow spreads. First, as Christie and Schultz (1994) emphasize, dealers seemed to coordinate their quotation patterns by only using price fractions that ended in even-eighths (for example a bid of $20.25 and an ask of $20.50), which meant that spreads were at least $0.25. Second, the practice of "preferencing" assured dealers they would receive order flow whether or not they quoted the best price. Preferencing is the practice by which brokers would route their customer orders to particular dealers (or their own dealer subsidiaries) in return for payments or other benefits with the stipulation that orders would execute at the best price in the marketplace. Since much of the order flow in a stock was preferenced, a given dealer could not attract much additional order flow by improving quotes. Instead, narrowing the quote would have the effect of cutting revenue on the existing business for all dealers (since each dealer committed to trade at the best quote). Consequently, the benefit to a dealer of improving the quote was low, and as a result, dealers limited their competition on quotes. Third, while quote competition was limited on the public Nasdaq market, dealers did offer competitive quotes over interdealer trading systems in order to manage their inventory. These quotes were not available to the general public, however. The net effect of these factors was to raise the bid-ask spread above the competitive level, especially for smaller trades. Huang and Stoll (1996), for example, show that spreads in Nasdaq stocks in the early 1990s were twice those of comparable NYSE stocks.

The controversy over Nasdaq trading procedures ended in 1997 with the Securities and Exchange Commission's promulgation of order handling rules that have transformed the Nasdaq Stock Market. One rule requires Nasdaq market-makers to display and make accessible customer limit orders, thereby enabling customers to compete directly with dealers to narrow the spread. For example, suppose the best dealer bid price is $20.25. Under the rule, customers could now enter superior quotes, say a bid of $20.375, to narrow the spread, something that had not been permitted before when only dealers could enter quotes. Furthermore, any customers selling to a dealer under preferencing arrangements (which promise the best price anywhere) would receive $20.375 rather than $20.25 so long as the limit order remained outstanding. (Of course, when the limit order traded, the bid might go back to $20.25, but new limit orders could be placed.) The second major rule was aimed at eliminating the practice by which dealers quoted better prices to each other than to the general public. The rule permits a dealer to enter better prices elsewhere only if they are posted in an electronic communications network (ECN) that makes its quotes generally available. The advent of ECNs, which are fully electronic and fully automated systems, may be ascribed to the 1997 order handling rules.

Electronic Communications Networks

The idea of electronic trading is not new. In 1971, Fischer Black suggested steps toward a fully automated exchange that would eliminate the need for spe-

cialists and market-makers. He noted that "a stock exchange can be embodied in a network of computers, and the costs of trading can be sharply reduced, without introducing any additional instability in stock prices, and without being unfair either to small investors or large investors" (Black, 1971, part II). He had in mind a world in which investors would interact with one another with little or no human intervention. That world is near, but Fischer Black (were he still alive) would be surprised at how long it has taken. The automation at ECNs has achieved what Fischer Black had in mind.

ECNs automate the trading process, but they do not change the fundamental functions that are carried out in trading. ECNs operate like the book shown in Table 2, where investors place limit orders. Other investors view the book on their computer screen, and if the displayed price suits them, they place an order to trade against a displayed price.[4] If enough orders are entered in the book by natural investors, trading can proceed among natural investors without the intervention of dealers.

The key problem for an ECN is to attract enough limit orders to assure liquidity. If there are insufficient natural investors placing limit orders, one solution is to pay for limit orders. For example, one ECN agrees to pay 2 cents per share for limit orders. A second solution is to convince dealers to make markets in the ECN book. ECNs have provisions for dealer participation, although dealers do not enjoy the special privileges that they have in traditional markets. A third solution is to agree to ship incoming orders to other markets when those markets offer better prices. In this way, investors are assured of the best price and therefore are willing to send orders to the ECN in the first place. Over time one would expect relatively few ECNs to attract most of the order flow, and this has occurred. The dominant ECNs are Archipelago and Instinet. Each has acquired other ECNs. Redibook was acquired by Archipelago, and Island merged with Instinet.

ECNs have a number of advantages. 1) They are automatic. Once an order is submitted, trade execution proceeds without human intervention according to price/time priority, unlike traditional markets, where orders might be held by dealers. 2) They are anonymous. The identity of traders is not revealed, which can be of importance to certain traders. 3) They are low cost. ECNs earn income by charging a fee to market orders of about 3 cents per share, while they pay for orders that supply liquidity. 4) They are fast. Execution and confirmation are electronic and occur in less than a second. 5) They can be programmed to offer complex orders. For example, ECNs can offer contingent trades where the price is adjusted for changes in index prices or in the prices of other stocks.

ECNs are best suited to active stocks and moderate sized trades. They are less suited to small inactive stocks, which may require dealer sponsorship to maintain liquidity. Large trades, which today are negotiated upstairs or handled by floor brokers, also may not be well suited to the ECN approach. But this may change.

[4] The Archipelago book can be accessed at ⟨http://www.tradearca.com/marketdata/book_info.asp⟩.

What floor brokers have done in the past to trade a large order slowly over the day is now done by upstairs brokers who trade slowly over different electronic systems.

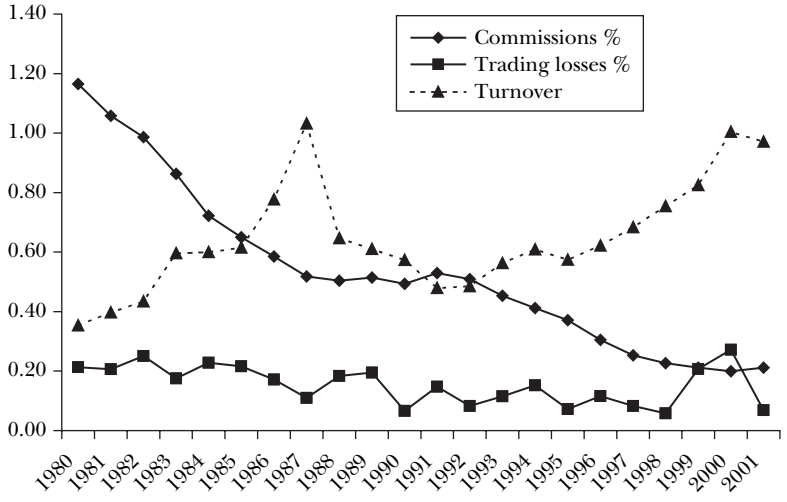
Aggregate Trading Costs

How have investor trading costs been affected by technological change and by regulatory change? Explicit trading costs are of two forms—commissions and trading losses. Commissions are paid to the broker for accepting and routing the order, clearing the trade, and maintaining the customer's account. Trading losses arise because investors pay the bid-ask spread to suppliers of liquidity and because trading may have an adverse price impact. Implicit trading costs are the costs of delay and the adverse price effects that investors may experience after an order is placed. Electronic markets have reduced all these costs. Commissions are lower because electronic systems have reduced the cost of handling orders. Trading losses and the bid-ask spreads are lower as dealer firms have consolidated and are realizing economies of scale. High-speed electronic trading makes it easier for customer limit orders to compete with dealers, thereby reducing the bid-ask spread. Implicit costs, such as delays, have been reduced with high speed electronic trading.

One way to measure the trend in the public's trading costs is to measure the trend in the aggregate revenues of securities firms. Commission revenues of securities firms are the commission costs of the public. Trading gains of securities firms are the trading losses of the public. Data for exchange listed stocks are provided in Figure 1. Commission revenues in listed stocks are reported to the Securities and Exchange Commission on brokerage firms' financial disclosure reports (FOCUS reports). Trading gains are also reported, and I estimated the portion attributable to listed stocks. The calculations follow Stoll (1995) and have been updated through 2001. Commission costs are stated as a percentage of public dollar volume. Since trades require both a buyer and seller, these figures reflect the cost of a round-trip (buy and sell). In 1980, the average round-trip commission was estimated at 1.17 percent. By 2001, this figure had declined to 0.21 percent. If stated as a percentage of share volume, the decline in round-trip commissions would be from 35.6 cents per share in 1980 to 6.6 cents per share in 2001. At the same time, trading losses due to the bid-ask spread and other market impact costs are estimated to have declined from 0.21 percent in 1980 to 0.07 percent in 2001 (interrupted by a surge in trading losses in the bubble years, 1999 and 2000). If stated as a percentage of share volume, the decline in trading losses would be from 6.5 cents per share to 2.14 cents per share.

As trading has become cheaper, turnover—annual share volume divided by shares outstanding—has grown. These data are also plotted in Figure 1. In 1980, turnover for exchange listed stocks was 0.35 per year. In 2000 and 2001, aggregate turnover of exchange listed stocks reached 1.0. (Turnover also temporarily surged to 1.0 in 1987, the year of the crash.) Those who believe that markets are domi-

Figure 1
Annual Percentage Round-Trip Trading Costs and Annual Turnover, Exchange Listed Stocks, 1980–2001.



nated by irrational noise traders view the reduction in trading costs and the rise in turnover with alarm, for it increases the chances that stock prices will be driven away from their fundamental values.[5] Others, who believe that lower trading costs and greater volume improve liquidity and allow rational traders to keep prices closer to their equilibrium value, applaud the increased trading efficiency of markets and the associated higher volume.

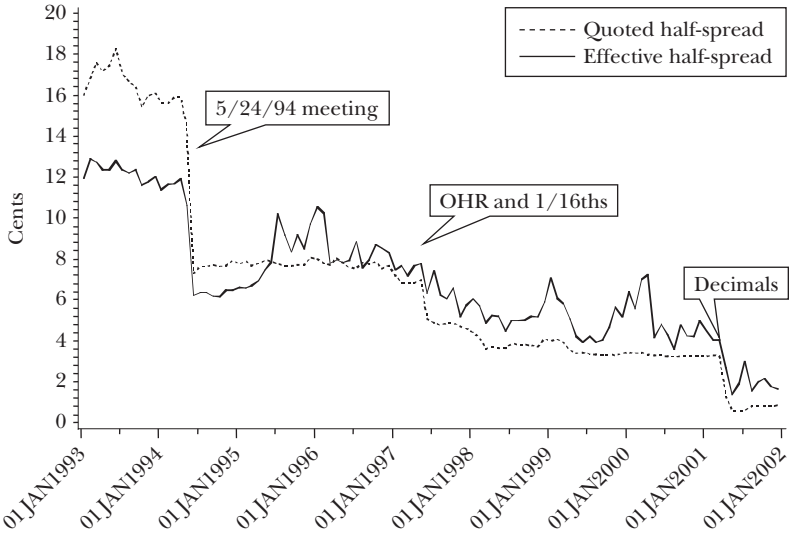
Whatever one's view on this debate, one point of agreement is likely to be that the securities firms have benefited despite the decline in trading costs. The increase in turnover combined with the increases in the shares outstanding has resulted in increased volume that has more than offset the decline in per dollar trading costs. In 1980, total commissions paid to trade exchange listed stocks were about $4 billion. By 2001, this number had grown in excess of $16 billion. Another indicator of the continued profitability of exchange membership is the price of a membership on the NYSE. In 1980, memberships (seats) sold for about $225,000. By 2001, seat prices had increased nearly 10-fold to about $2,150,000.

Bid-Ask Spread

Trading costs reported in the previous section are based on aggregate revenues reported by securities firms. Another gauge of investors' trading costs is the

[5] See, for example, Shleifer and Summers (1990) for the noise trader approach to finance.

Figure 2
MSFT 1993–2001, Daily Average Quoted and Effective Half-Spread



bid-ask spread. Changes in bid-ask spreads over the past 10 years have been dramatic. Figure 2 plots the daily average quoted and effective half-spreads for Microsoft (MSFT), which is representative of heavily traded Nasdaq stocks. Figure 3 provides the same data for IBM, which is representative of heavily traded listed stocks. We plot the half-spread, which reflects the cost of one trade, rather than the spread, which reflects the cost of a round-trip of two trades.

The quoted half-spread is half the difference between the best bid and best ask. The effective half-spread is the difference between the transaction price and the midpoint of the quotes. If the transaction is inside the spread, the effective half-spread will be less than the quoted half-spread. For each day, the half-spreads associated with each trade are averaged over all trades in the day, and the monthly average is plotted in Figures 2 and 3.

In 1993, the quoted half-spread in Microsoft exceeded 16 cents. By the end of 2001, the quoted half-spread had declined to under 2 cents. On Nasdaq, the most dramatic decline occurred as a result of a May 1994 meeting of the Nasdaq administration with dealers to urge a reduction in spreads. The meeting was called in response to the Christie and Schultz (1994) paper and the emerging controversy over Nasdaq dealer behavior. Subsequent declines occurred with the order handling rules and reductions in the tick size. Prior to 1997, the minimum tick size was one-eighth dollar, or 12.5 cents, on both the NYSE and Nasdaq. This minimum necessarily resulted in a minimum spread of 12.5 cents. Regulatory action reduced the tick size to 6.25 cents in June 1997 and to 1 cent in 2001. In 1993, IBM's quoted

Figure 3
IBM 1993–2001, Daily Average Quoted and Effective Half-Spread

half-spread averaged about 8 cents. By the end of 2001, IBM's quoted half-spread had fallen to about 2 cents, as shown in Figure 3.

The decline in spreads primarily reflects regulatory action (especially vis-a-vis Nasdaq) and the reduction in tick size, but electronic trading has been a continuing factor in reducing trading costs. Electronic trading technology makes it easier for dealers to manage large inventories and thereby to manage their risks more effectively. Technology has helped make consolidation of dealer firms possible. As dealers have merged, their stock portfolios have become increasingly diversified, which has reduced their risk, and at the same time made possible a more flexible allocation of market-making capital.

A noteworthy feature of the Microsoft graph is the trend in the effective half-spread relative to the quoted half-spread. In 1993, the average daily effective half-spread was less than the quoted spread. This means that, on average, trades took place inside the spread—customers were able to negotiate prices superior to the posted quotes. After 1996, average daily effective half-spreads exceeded average daily quoted half-spreads. This means that, on average, trades took place outside the spread. Since investors would normally prefer to trade at the spread rather than outside, the presence of trades outside the spread reflects inefficiencies in the market mechanism. This does not mean that the market is less efficient after 1996 than before—only that the wide quoted spreads before 1996 hid the inefficiencies. If markets were fully electronic with immediate automatic execution, one would not observe any trades outside the quotes. The fact that outside trades are observed implies that some trades failed to receive the best price or that trades were reported

with delay after the quote had changed. In addition, trades outside the quotes reflect the fact that large trades cannot be traded entirely at the prevailing quote.[6] A large order may decide to ignore small better-priced orders so that the entire amount of the order can be quickly traded. Almost no outside trades are observed on the NYSE. This is because the market is more centralized with one dealer, whereas the market for Microsoft is dispersed among many dealers who may be quoting different prices.

Dealers versus Customers

The decline in the bid-ask spread depicted in Figures 2 and 3 reflects the increased efficiency in trading, the move to decimal pricing and greater competition, yet the fact remains that dealers earn revenues from customers. Trading costs have declined, but the data on spreads may overstate that decline. As the figures show, the quoted half-spread has declined with decimalization. Yet there is not complete agreement that decimal price increments have reduced the overall cost of trading. This is because the depth at the best quotes has also declined. An investor who wants to buy 1,000 shares will pay more than the quoted ask if the amount available at the ask is only 500 shares. Bacidore, Battalio and Jennings (2003) find that higher prices must be paid for larger quantities with the result that execution costs are unchanged after decimalization.

A subtle issue in the trading process is the fact that trading remains a negotiation in which customers can find themselves at a disadvantage to dealers. In negotiating a trade, one side proposes a transaction, for example, by entering a bid or placing a limit order, and the other side responds by accepting the bid or trading against the limit order. Consider a limit order to buy at $20 when the current ask price is $20.10. The limit order is placed in hopes of buying the stock at $20 rather than paying the ask price of $20.10, but the limit order faces an asymmetric risk. If bad news justifies a price for the stock of $19, someone will sell to the standing limit order at $20, and gain $1 at the expense of the limit order, a loss not reflected fully in the spread. On the other hand if the news is good, someone is likely to better the price of the limit order so that it does not execute. Because the minimum tick size is now $0.01, the cost of bettering the $20 bid is small—only one cent. The process is in fact termed "pennying." The dealer (or another speedy investor) need only quote $20.01 to "penny" the standing order. In effect the limit order has granted a put option—a floor—to the rest of the market to sell to it at $20 while at the same time giving other traders the opportunity to better the price if conditions warrant. The winner of this game is, according to an old Wall Street adage, "he who trades last." The dealer or other professional trader is more likely to be in a position to

[6] See Stoll and Schenzler (2005) for evidence on the nature and sources of outside trades.

exercise the option—to have the last say—than the public investor, even in electronic markets.

Electronic trading is not the same as automated trading. Even in electronic markets, subtle trading costs arise when one side of the trade can introduce delays and thereby create valuable trading options. Suppose the dealer's ask price is $20.10. If order routing and execution were automatic, a market order to buy would trade at $20.10. However, execution need not be automatic. For example, although most orders arrive electronically and are displayed on the dealer's screen, execution requires action by the dealer, which takes time. During that time, which need not be longer than 15 seconds in an active stock, the dealer has an option of how to trade the market order. That option has value to the dealer and is a cost to the customer. For example, suppose that in the 15 seconds the dealer learns that sell orders are arriving. Knowing that, the dealer will be happy to sell to the existing market order at the initial price of $20.10 and cover this sale by buying from the incoming sell orders at lower prices. If, on the other hand, the dealer learns that more buy orders are arriving, he will raise the ask price above $20.10 or let the market order trade against a limit order. Delay creates valuable flexibility for the dealer. When trading is automated, delays cannot be introduced and trading options do not arise.

Electronic trading, even if not fully automated, constrains the negotiation process and creates an audit trail that benefits public investors. Twenty-five years ago, it might have taken an hour to complete a trade. During that hour, brokers and dealers could observe the order and would have options in how to trade with the order. Today, trades are completed far more rapidly, and the process is more transparent. It is easier to monitor an order now than before. Negotiation may still take place, but it is in a much tighter time frame. As trading becomes fully automatic, the discretion of the dealer, discussed above, is eliminated, and problems of the type discovered by the recent Securities and Exchange Commission investigation of specialists will be reduced.

New electronic trading systems are strengthening the hand of customers vis-a-vis dealers. Limit orders can be programmed to change their limit prices when market conditions change, so limit orders can more effectively compete with dealer quotes. Market orders can be programmed against execution delay to cancel if they are still outstanding after 15 seconds, thereby limiting the option of the dealer who may be holding the order. As execution is automated, customers benefit because the dealer's option to delay execution of orders is eliminated.

Computer Trading

To what extent can computers take over stock trading? Is it possible that the future stock exchange will be a battleground in which the computers of different investors face off without the intervention of a person? I believe the answer to this question is "yes, in part." Computers now play an important role as tools in routing

and executing trades. The decisions of what to trade are, however, for the most part made by humans. But one can imagine a future in which stock selection programs link directly to trading programs.

One type of computer trading that is common today is program trading. The NYSE defines a program trade as a trade involving 15 or more different stocks totaling at least $1,000,000 in value. Some trades might involve all the stocks in the Standard & Poor's 500 Index. The term "program" is used because computers generate the trading instructions in the individual stocks. Suppose an institutional investor wishes to buy $25 million of the S&P 500 Index. Instead of placing 500 different orders, the investor places one order, and the computer sends orders for the appropriate number of shares of each of the 500 stocks in the index. The computer is not a decision maker, but a trade facilitator. According to the NYSE, program trading currently accounts for roughly 25 percent of aggregate NYSE volume, up from 13.9 percent as recently as 2001.[7] Program trading reflects the increasing tendency to hold and therefore to trade standard portfolios. Since many mutual funds replicate a particular index, it is logical that trading would take the form of trading the index portfolio.

One use of program trading is to accomplish index arbitrage between the current value of the Standard & Poor's 500 Index and the futures contract on the index. Index arbitragers purchase the S&P 500 stocks and sell index futures if the futures price is above its fair value and, conversely, sell the S&P 500 stocks and purchase index futures if the futures price is below its fair value. Index arbitrage was an important reason for program trading in the 1980s, but currently it accounts for only about 6 percent of program trading. Most program trading is used to accomplish broader portfolio objectives—to shift into or out of sectors of the market, to invest new funds efficiently, to liquidate a portfolio or to make a transition from one portfolio to another.

A second form of computer trading is smart order routing. A smart router is a computer program that sends an order to the best market. The best market is defined by price, speed of execution and other characteristics of the market. Thus, for a stock like IBM, which can be traded on the NYSE, Boston, an ECN and Nasdaq, a smart router will determine which market has the best price and has in the past been quickest to trade and send the order to that market. The computer may also manage the order, for example, by canceling it if it has not executed within 15 seconds, by adjusting the limit price if the order is a limit order or by sending orders to test the depth of a market. If the order is a large order, the process of trading can be complex. The order has to seek markets with sufficient liquidity, which is not always displayed on the limit order book (because traders do not want to show their hand), and the order may have to be split into pieces. Computers can be programmed with algorithms that accomplish much of what a

[7] See ⟨http://www.nyse.com/Frameset.html?displayPage=/marketinfo/ProgramTrading.html⟩. The statistic means that program purchases are 25 percent of trading volume and that program sales are 25 percent of trading volume.

trader might do in handling a large order. For example, an investor might instruct the computer to buy at the average price for the day, which the computer would accomplish by trading throughout the day.

A third form of computer involvement is in making actual investment decisions. For example, quantitative portfolio managers use computers to identify underpriced and overpriced stocks. In effect, the computer makes the investment decision based on a model and on the information made available to it. The next step would be to integrate the stock selection decision and the trading decision so that the entire process is computerized. While this third form of computer trading is not yet common (except for well-defined strategies such as arbitrage, or short term trading strategies), it is likely to gain in use. Does this mean that human judgment will disappear from investment decision making? Not likely. Ultimately, the computer must be instructed. In the end, successful investment management is likely to require human input and human ingenuity.

Should we fear computer trading in any of its forms? What can go wrong? Mistakes in order entry could be more serious with computer order routing. If a clerk enters "sell 1,000,000 shares" when the sale should have been for 1,000 shares, and if the computer blindly carries out the instructions, market prices will be dramatically affected. In a human market, such an error might be caught by another clerk or a floor broker, but in a computer market, no one might catch the error until trading has taken place. On Friday, December 5, 2003, such an incident occurred when a computer program mistakenly entered thousands of trades in Corinthian Colleges, a Nasdaq stock. Between 10:46 AM and 10:58 AM, when trading was halted, approximately a net quantity of 3,000,000 shares was mistakenly sold, which drove the price from $57.50 to $39.50. By the end of the day, the stock recovered to $56.80 (Norris, 2003). Computer failures, either at the level of the broker who enters orders, or at the level of markets receiving orders, can disrupt trading. If the computer goes down, trading on one or more markets may be halted.

Another concern is that computers may be used to game prices. For example, smart routers may place temporary fake buy limit orders in the hopes of pushing prices up and then sell shares at the higher price. One smart computer may attempt to take advantage of the order placement strategy of another not-so-smart computer with the possible effect that prices are driven away from equilibrium. Two forces limit gaming. First, investors will be wary of gaming by other investors. Second, outright manipulation is against the law. Computer audit trails enable regulators to monitor such behavior.

Market Structure: Consolidation versus Fragmentation

Electronic trading is having a major effect on the structure of securities markets, that is, on the number and types of exchanges and how they interact.

While important forces are leading to consolidation of markets, many observers believe stock trading is excessively fragmented.

Two economic forces tend to lead to the consolidation of trading in a single market—economies of scale in processing orders and network externalities—but the centralizing tendency can be offset if communication across markets is easy and if orders can readily be routed from one market to another. Economies of scale exist on the production side because the average cost of handling an order declines with the number of orders. In effect, the fixed costs of an exchange can be amortized over more trades. Hence, in the absence of offsetting factors, an exchange is a natural monopoly. Because an existing exchange can usually underprice a new entrant, it will have a first mover advantage.

Network externalities arise on the demand side because trades are easier to accomplish if others are trading at the same location. A centralized market maximizes the possibility of interaction with other traders—hence, traders want to trade where others are trading. The tendency toward consolidation is particularly strong if the only form of communication is face-to-face. When the primary form of communication is face-to-face, economies of scale and network externalities tend to lead to one primary location for trading in each security. In the past, the NYSE was that location for the major stocks. The AMSE and Nasdaq have been able to carve out market segments where they are the leaders.

The natural monopoly of an exchange is threatened by advanced communications. For example, with the advent of the telephone, competitors could telephone the NYSE for current prices and offer to trade at those prices at a lower commission. This is how the American Stock Exchange (once known as the curb market) got its start, and it is the way in which regional exchanges have maintained their existence. To limit such "free riding," the NYSE prohibited phone contact between the floor and nonmembers off the floor, but that limitation was ruled illegal by the courts in *Silver v. New York Stock Exchange* (373 U.S. 341 [1963]). Since then, improvements in technology and regulatory intervention have increased communication with, and access to, the NYSE. Increased communication with the primary market makes it easier for a new market to compete because it can demonstrate that its prices match or improve on those of the primary market. Thus, with improved communication, the forces to fragment trading are stronger today and the forces of consolidation are weaker. Some view this as desirable because competition is increased; others view it as undesirable fragmentation of liquidity.

Regulation and Market Structure

Regulation has affected the balance between consolidation and fragmentation of markets. The Securities Acts Amendments of 1975 authorized the Securities and Exchange Commission to facilitate a national market system (NMS) with fair competition among brokers, dealers and markets, the availability to all of information on transactions prices and dealer quotes, the linking of markets and the ability

to execute orders in the best market. Some interpreted the 1975 amendments as calling for the Securities and Exchange Commission to develop a single integrated national market system in which all orders would be routed to the national consolidated limit order book (CLOB). The Securities and Exchange Commission has vacillated over the years in the degree of forcefulness with which it pursued a national market system. On the one hand, a single national market system with a single consolidated limit order book would inhibit the innovation and competition that new markets and new trading systems bring. On the other hand, allowing too many independent markets would, the Securities and Exchange Commission feared, lead to inefficient price formation and to fragmentation of the price discovery process. As a result the Securities and Exchange Commission has tried to accommodate new markets, such as the electronic communications networks, while improving links among markets.

Recently, the Securities and Exchange Commission approved the latest in a series of rules to link markets while at the same time allowing for independent development by each market. On April 6, 2005, the Securities and Exchange Commission, under its authority (Section 11A of the Securities Exchange Act of 1934) to facilitate the establishment of a national market system, approved a controversial rule—the order protection rule (also known as the order trade-through rule) that requires a market receiving an order to send the order to any other market that has better posted prices if those prices are automated and immediately accessible (Securities and Exchange Commission, 2005). Put differently, the rule prohibits an incoming market order from ignoring (trading through) a better quote in market A to trade with a poorer quote in market B. Quotations that are not immediately accessible are not protected under the rule.

The rule is a big change for the Nasdaq Stock Market in which an order protection rule has been absent. An order protection rule has applied to NYSE-listed stocks across the markets in which listed stocks trade, but that rule has failed to operate effectively because it applied to quotes that were not immediately accessible. For example, it was frequently the case that an order sent to the NYSE to trade with an apparently better quote was not executed because the NYSE specialist had the opportunity to withdraw the quote and update it manually. The effect was that high-speed electronic orders were compelled to wait for manual quote updates. The new rule protects only those orders that may be executed automatically and without delay. This requirement puts pressure on the NYSE to automate its trade execution process.

The rule is controversial for several reasons. Opponents argue that the rule interferes with the right of customers (through their brokers) to send orders to any market and improperly injects the government into everyday business decisions. They argue that customers should have the right to base order routing on factors other than price, such as reliability, speed and creditworthiness. Consumers are not required to buy their TVs where the price is lowest, so investors should not be required to buy their stocks where the price is lowest.

Supporters of the rule argue that price priority is central to a well-functioning

market—the rule is only doing what brokers, acting in their customers' best interests, should do. They are concerned that brokers may not act in their customers' best interests and may preference orders to markets with which they have special deals that may not offer the best price. Other observers argue the rule has not gone far enough because it applies only to the "top of the book," namely the best bid and offer in a market. Second-best quotations are not guaranteed price priority by the rule. Consequently, a market that bid $20.01 and $20.00, when the best bid in all other markets was $19.99, would have its bid of $20.01 protected by the rule, but not its bid of $20.00. For example, suppose market A quotes $20.01 for 600 shares and $20.00 for 300 shares, and market B quotes $19.99 for 300 shares. A market sell order for 900 shares would be obligated to trade at the best "top of the book" prices, which would be to sell 600 shares in market A at $20.01 and 300 shares in market B at $19.99.

The casual observer of the heated debate that has surrounded the order protection rule may well wonder what the fuss is all about. After all, we are just talking about pennies. But for the exchanges, it may be a matter of business survival. Pennies matter, but more important, the rule requires the linkage of markets, which threatens established markets and benefits new markets. The battle appears to be over pennies, but in fact, it is over the ability of markets to separate themselves from the pack. The linkage required by the Securities and Exchange Commission links trading centers directly. Each trading center, such as an electronic communications network, would be linked to every other trading center, such as the NYSE, and would be required to send orders to the competing market if that market had better quotes. A direct linkage tends to limit competition among markets because it forces markets to integrate their operations; in fact, it requires an exchange to send orders to a competitor. (It is like asking Wal-Mart and Target to cooperate by sending each other customers.) But market centers need not be linked directly. The ability of brokers to send orders electronically to any market indirectly links markets. Instead of requiring each market to assure that orders on all other markets are protected, an alternative approach would be to let the broker handling the customer's order determine the best market. The broker can create a virtual electronic book that contains the best quotes and quantities from every market and can send orders directly to the best market. Under this approach, markets would be linked by the routing decisions of brokers rather than at the level of the market center.

Conclusions

Stock markets are central to modern capitalist economies. The prices formed on them provide signals about the economic viability of firms and facilitate the proper allocation of resources among firms. Markets provide liquidity to investors that wish to save or consume. Electronic trading has improved the efficiency of stock markets and hence has reduced the cost of

providing liquidity and has increased the accuracy of price signals. Electronic trading benefits investors by increasing the speed and lowering the cost of trading while at the same time creating a complete audit trail that facilitates monitoring of brokers. In electronic markets, there is less opportunity for professional traders to take advantage of public investors. One worry is that dependence on computers subjects the markets to computer glitches or errors that might be avoided in a manual market. Another is that the low cost of trading will lead to excessive uninformed trading.

An important question for the future is how electronic trading will affect the structure of the stock market. Will the NYSE survive as a floor based market? How many electronic communications networks will survive? These questions are rapidly being answered as this article is being written. After agreeing to purchase the Pacific Stock Exchange in January 2005, the ECN, Archipelago, agreed to merge with the NYSE in April 2005. Simultaneously, the other major ECN, Instinet, owned by Reuters, was sold to the Nasdaq Stock Market. These events suggest that the economies of scale of the trading business are pushing markets to consolidate. At the same time, dealer firms are consolidating, as is evidenced by the decline in the number of specialist firms on the NYSE. The outlook is thus for fewer markets and fewer dealers. To stay competitive, the NYSE is likely to become more like an ECN, which means that it will ultimately give up the trading floor as a burdensome expense.

The advent of automated trading does not mean that the function of dealers, brokers and specialists will disappear, only that the current way in which these functions are carried out will change. Even in automated markets, there is likely to be a need for liquidity provision, particularly in less active stocks. Professional traders, not necessarily called "specialists" or "market-makers" and not necessarily having an affirmative obligation, will provide liquidity because it is a profitable activity. At the same time, the privileges enjoyed by today's designated market-makers (such as information about pending orders) are likely to be eliminated. The form in which floor brokers operate is also likely to change: instead of trading on the floor on behalf of their institutional customers, they will do their trading over computer screens located in upstairs offices. In the end, venerable institutions, like the NYSE, will be different, although the functions they provide will remain.

■ *The helpful comments of Bill Christie, James Hines, Christoph Schenzler, Andrei Shleifer, Rene Stulz and Timothy Taylor are gratefully acknowledged.*

References

Anderson, Jenny. 2005. "15 Specialists from Big Board are Indicted." *New York Times.* April 13, p. C1.

Bacidore, Jeffrey, Robert Battalio and Robert Jennings. 2003. "Order Submission Strategies, Liquidity Supply, and Trading in Pennies on the New York Stock Exchange." *Journal of Financial Markets.* 6:3, pp. 337–62.

Black, Fischer. 1971. "Toward a Fully Automated Exchange." *Financial Analysts Journal.* July/August and November/December, 27, pp. 29–35 and pp. 86–87.

Ceron, Gaston F. 2005. "Moving the Market: NYSE Seat Sale Hits $3 Million." *Wall Street Journal.* August 4, p. C3.

Christie, William G. and Paul H. Schultz. 1994. "Why do NASDAQ Market-Makers Avoid Odd-Eighth Quotes?" *Journal of Finance.* 49:5, pp. 1813–840.

Christie, William G. and Paul H. Schultz. 1995. "Did Nasdaq Market-Makers Implicitly Collude?" *Journal of Economic Perspectives.* 9:3, pp. 199–208.

Christie, William G., Jeffrey H. Harris and Paul H. Schultz. 1994. "Why did NASDAQ Market-Makers Stop Avoiding Odd-Eighth Quotes?" *Journal of Finance.* 49:5, pp. 1841–860.

Demsetz, Harold. 1968. "The Cost of Transacting." *Quarterly Journal of Economics.* February, 82, pp. 33–53.

Huang, Roger D. and Hans R. Stoll. 1996. "Dealer versus Auction Markets: A Paired Comparison of Execution Costs on NASDAQ and the NYSE." *Journal of Financial Economics.* 41:3, pp. 313–57.

Norris, Floyd. 2003. "Nasdaq and the Pacific Exchange Criticize Each Other's Actions during a Computer-Related Mishap." *New York Times.* December 9, p. C11.

Securities and Exchange Commission. 2004. "Press Release: Settlement Reached with Five Specialist Firms for Violating Federal Securities Laws and NYSE Regulations." March 30; Available at ⟨http://www.sec.gov/news/press/2004-42.htm⟩.

Securities and Exchange Commission. 2005. "Press Release: SEC Adopts Regulation NMS and Provisions Regarding Investment Advisers Act of 1940." April 7; Available at ⟨http://www.sec.gov/news/press/2005-48.htm⟩.

Shleifer, Andrei and Lawrence Summers. 1990. "The Noise Trader Approach to Finance." *Journal of Economic Perspectives.* 4:2, pp. 19–33.

Sofianos, George and Ingrid Werner. 2000. "The Trades of NYSE Floor Brokers." *Journal of Financial Markets.* 3, pp. 139–76.

Stoll, Hans R. 1985. *The Stock Exchange Specialist System: An Economic Analysis. Monograph Series in Finance and Economics.* Monograph 1985-2. New York, N.Y. : New York University.

Stoll, Hans R. 1995. "The Importance of Equity Trading Costs: Evidence from Securities Firms' Revenues," in *Global Equity Markets: Technological, Competitive, and Regulatory Challenges.* R. Schwartz, ed. Homewood, Ill.: Irwin, pp. 98–120.

Stoll, Hans R. 1998. "Reconsidering the Affirmative Obligation of Market-Makers." *Financial Analysts Journal.* 54:5, pp. 72–82.

Stoll, Hans R. and Christoph Schenzler. 2005. "Trades Outside the Quotes: Reporting Delay, Trading Option, or Trade Size?" *Journal of Financial Economics.* Forthcoming.