

The Structure of the U.S. Equity Markets

Marshall E. Blume¹

The Wharton School, The University of Pennsylvania

Presented at

The Federal Reserve Bank of Atlanta

2000 Financial Markets Conference

Sea Island, Georgia

October 15–17, 2000

In 1975 Congress directed the Securities and Exchange Commission to develop a national market system in which all orders to buy or sell equities would interact. A national market system abhors fragmentation and assumes that one market will best serve the needs of all investors. Such an assumption does not capture the realities of modern markets. Investors have different needs, and different markets will develop to serve these needs. Fragmented markets are a natural result of competition. Within the United States, the sharing of trade and quote information among markets helps to mitigate any deleterious effects of fragmentation. The markets of tomorrow will be global. In a global market, the Securities and Exchange Commission will have to give up its goal of a national market system and focus on other issues. For example, it will be a challenge simply to provide the sharing of trade information across borders. Further, technology will allow a market center or order-gathering function to be located anywhere in the world. This threat of relocation will place constraints on U.S. regulators, and global trading will make it more difficult for U.S. authorities to regulate investment practices and protect U.S. investors.

¹ Comments and advice were received from Robert Eisenbeis, Dean Furbush, Adam D. Goldberg, Michael L. Goldstein, Richard J. Herring, Kenneth Kavajecz, Bernard Madoff, Peter Madoff, Tim McCormick, John G. Moran, Gus Sauter, Jeremy Siegel, James Shapiro, and George Sofianos. The author thanks them for their comments; however, none of the views expressed in this paper should be attributed to these individuals.

The Structure of the U.S. Equity Markets

Equity markets worldwide are in a state of change, and technology and the Internet have and will continue to have a profound impact on their structures. Recent Securities and Exchange Commission (SEC) rulings have unleashed competitive forces in the U.S. equity markets that may result in new exchanges to challenge the established markets.

To date, the cost and awkwardness of the clearing and settlement processes across nations have restrained global competition, but these processes are changing. As cross-border settlement becomes cheaper and easier, global trading may increase and U.S. regulators will need to adapt to this new environment.

The 1975 Amendments to the Securities Exchange Act of 1934 (SEA 1934, section 11A) set as a national goal that all securities should be traded in a national market system. This goal has shaped much of the SEC's thinking since that time, but it may collide with some of the ongoing structural changes occurring in the equity markets, both domestically and worldwide. As argued below, a national market system represents a naive and parochial view of the way equities are traded. Such a system is incompatible with how markets currently behave and how they will develop in the future.

A national market system abhors fragmentation because fragmentation limits interaction among order flow. Yet fragmentation is at the heart of competition. New competition creates fragmentation, but significant fragmentation will occur only if the competition is successful. Extremely successful competition may well drive out existing markets, resulting ultimately in less fragmentation.

To be effective in this increasingly global world, the SEC must give up its goal of a national market system, begin to recognize the global nature of the equity markets, and shape its regulatory thrust to the regulatory issues of a global market. The SEC and other U.S. regulatory authorities will have to coordinate their activities with those of other nations.

In future years the stock of a firm like Microsoft will trade worldwide, at least for retail customers. Investors in Asia will be able to buy U.S. stocks during their normal business hours in their local currency; the same will be true for investors in Africa, Australia, Europe, and South America. Of great importance to U.S. regulators is that a U.S. investor will be able to trade U.S. stocks in foreign markets. By so trading, U.S. investors may be able to both legitimately

circumvent some U.S. regulations and illegitimately hide questionable activities from U.S. surveillance.

The movement toward global trading has begun and will continue. When global trading becomes a reality, the country and currency in which an investor trades will be a matter of choice. With just a click, a U.S. investor will be able to trade on any market that provides advantages over U.S. markets. Still, it is quite possible that the bulk of trading in a security will remain in the domestic market. To be successful a market requires a critical mass of liquidity. Since the domestic market already has that mass, it is difficult, though not impossible, for another market to establish itself.

Monitoring and regulating this movement to global trading will be one of the major regulatory challenges over the next decade. It will require that domestic regulators coordinate their regulations with those of other countries.

A nonexhaustive list of challenges in regulating worldwide trading includes overseeing a global clearing and settlement system, preventing fraud, maintaining transparency, enforcing stockholder rights, collecting taxes, and integrating disparate and perhaps contradictory laws.

The next two sections of this paper describe the legislated national market system and its limitations. Since the concept of a national market system underlies much of the current U.S. regulatory efforts, it is important to understand this concept and its limitations. The following sections highlight the current organization of the markets, including the consolidated reporting of information, participants, and regulations. The final sections examine some of the major issues of today and argue that these issues will change as trading becomes increasingly global. The paper concludes with some thoughts about the future.

The Legislated National Market System

The Securities Acts Amendments of 1975 (SEA 1934, section 11A) directed the SEC to establish a national market system. This directive has shaped much of the regulatory efforts of the SEC since that time.

To quote the amendments, “The Commission is directed . . . to use its Authority . . . to facilitate the establishment of a national market system for securities” (SEA 1934, section 11A[a][2]). The amendments are less clear as to what constitutes a national market system. However, the amendments observe that “the linking of all markets for qualified securities

through communication and data processing facilities will foster efficiency, enhance competition, increase the information available to brokers, dealers, and investors, facilitate the offsetting of investors' orders, and contribute to the best execution of such orders" (SEA 1934, section 11A[a][1][D]).

Thus, the concept of the national market system involves a link of some type among the various markets and certainly includes the idea of price priority over all markets. Price priority requires that no market execute a trade at a worse price than that available on another market. For example, if one market posts an offer to sell ABC at \$20, no investor should pay more than \$20 in another market.

The amendments were silent on the issue of time priority within price. Time priority requires that in the presence of two or more offers to sell at the same price, the offers be executed in the time order in which they were submitted. Similar time priority applies to bids.

At the time the amendment passed, some observers, including Morris Mendelson, R.T. Williams, and Jay Peake, took it for granted that a national market system would include time priority. In response to an SEC request for proposals, they outlined in 1976 the major components of a consolidated limit order book, or CLOB. A CLOB works as follows: Some investors submit limit orders to a centralized book. The orders in this book are executed according to price-time priority. Other investors submit market orders to be executed against these limit orders. To provide strict price-time priority, all trades must take place through this CLOB; no orders can be executed outside the CLOB. In their model, all orders are anonymous.

On the surface, this system appears to be eminently fair. Any investor who submitted a limit order would be assured that his or her order would be executed according to price-time priority. Any investor who submitted a market order would receive the best price available at the time.² No one would have an unfair advantage.

This model of interaction still governs SEC thinking. In its recent concept release on market fragmentation, the SEC requested comments on the desirability of requiring an electronic integration of all limit orders for a specific security with the goal of preserving price-time

² It is not generally pointed out that in a CLOB market orders themselves would never cross within the best bid and offer as there is zero probability that two market orders would arrive simultaneously in such an electronic market. The crossing of two market orders within the bid and offer as occurs on the NYSE happens because market orders are often not executed immediately.

priority across all markets.³ This proposal echoes the early discussion of a centralized limit order book with the important exception that certain orders would be exempted, such as institutional orders crossed in the upstairs markets. The effect of this exception is that some orders would be allowed to trade ahead of limit orders in the central book, an effect that some would interpret as contrary to the intent of the 1975 amendments.

Interestingly, the subsequent release on order routing disclosure reported that “many commenters [on the release on market fragmentation], especially institutional investors, supported a nationwide system of price/time priority.” Further, these commenters expressed “serious concern about market fragmentation.”⁴

The Fallacy of “One Market Fits All”

Underlying the call for a national market system is the proposition that there is one best price at any point in time and that price is best achieved in one type of market structure. In this view fairness demands that an investor submitting a market order receive this best price. The following discussion will argue that the very concept of a best price is not well defined and investors might rationally prefer different types of markets. Even the concept of fairness can be defined only within a specific market structure.

Further, competitive markets will become fragmented in response to the diverse demands of investors. Some investors will prefer one type of market while others will prefer other types. No single market structure will satisfy the needs of all investors. Some fragmentation is a natural result of competition.

Net versus Gross Price. The debate about best price has virtually always centered on the best net price, that is, before commissions. In the trading of most New York Stock Exchange-listed stocks, the broker typically adds a commission. In the trading of Nasdaq-listed stocks, an investor often pays just the net price; however, investors today, particularly retail investors, increasingly are paying a commission in addition. In the end, the price that an investor pays or receives is the gross price, not the net price.

To state the obvious, the price of ultimate concern to an investor is the gross price, not the net price. In the traditional model where an order gatherer receives an order and then sends it

³ See SEC Release No. 34-42450, File No. SR-NYSE-99-48 <<http://www.sec.gov/rules/sro/ny9948/sprague1.txt>>.

⁴ See SEC Release No. 34-43084, File No. S7-16-00 <<http://www.sec.gov/rules/proposed/34-43084.htm>>.

to a different market for execution, the net price and the commission are uniquely determined because each is determined in a separate market. However, in a vertically integrated firm that acts both as an order gatherer and as market maker, such a distinction between net price and gross price becomes blurred. For any gross price, the net price is arbitrary since the commissions can always be adjusted upward or downward as necessary.

In today's world the New York Stock Exchange (NYSE) asserts correctly that a market order submitted to it is often executed within the national best bid and offer (NBBO) when the spread is two or more ticks. Thus, by submitting an order to the NYSE, an investor might obtain what is termed "price improvement." But an investor might rationally prefer sending an order to another market that does not offer "price improvement" if the gross price from using that market is less than the gross price from using the NYSE.

But the issue is more complicated. To obtain price improvement on the NYSE, there can sometimes be a delay in execution of thirty seconds and occasionally more. Other markets, which might not provide as much price improvement as the NYSE, may provide quicker execution. A retail customer might choose to forgo the possibility of obtaining a better net price, or even a better gross price, in return for quicker execution. Indeed, a survey conducted by Sanford C. Bernstein & Company found that 58 percent of on-line traders rate immediacy of execution as more important than a favorable price in evaluating the quality of a trade execution (Bernstein Research 2000b).

Different Needs. Two assumptions are implicit in the desirability of a national market system. The first assumption is that all investors would prefer such a market. Yet investors have different needs. As suggested above, some investors place a high premium on speed of execution, but the needs of investors differ much more than this simple example indicates. Some investors, and some major ones at that, prefer a nonanonymous market and will try to trade only with investors whom they know. These investors would not want a national market system, the very essence of which is anonymity.

A second way to justify a national market system is to assert that such a system involves externalities that would greatly enhance national wealth. Such externalities might include enhanced liquidity that could lead to a reduced cost of capital. This paper will not address this issue because the United States does not currently have a national market system and such externalities have not yet been established.

At the time of the call for a national market system, there was little recognition that some investors possess unrevealed information. Bagehot (1972) was perhaps the first to recognize that traders possess such information and to analyze how such unrevealed information would impact market prices. He assumed an anonymous market structure much like that envisioned by the national market system. Glosten and Milgrom (1985) formalized Bagehot's insight.

The effects are very simple but profound. Glosten and Milgrom proposed an anonymous market in which a market maker posts a bid for one share (a price at which it is willing to buy) and an offer for one share (a price at which it is willing to sell). The market maker acts competitively and thus sets the bid and offer at levels such that its expected profit is zero. The market maker faces no explicit costs such as fees, clerical expenses, and so on. In this model the spread arises solely to protect the market maker against losses to knowledgeable traders, not to cover the direct costs of being a dealer.

After the market maker posts a bid and an offer, an investor arrives and decides whether to hit either the bid or the offer. There are two types of investors: uninformed traders who need to trade for reasons of liquidity and informed investors who know something about the true price of the security. Although more restrictive than necessary, the following model assumes that the informed trader knows the true value of the security to which the market price will move in the near future.

This model assumes that a liquidity trader will always trade and an informed trader will trade only if it is to his advantage. To illustrate the decision of an informed trader, assume that the bid is 20 and the offer is $20\frac{1}{4}$. If the informed trader knows that the price will move to 19, he will certainly sell the stock at the bid of 20. If he knows it will move to $20\frac{1}{8}$, he will not trade. If he knows that the price will move to 21, he will buy at $20\frac{1}{4}$.

The market maker always loses to an informed trader. To break even on average, the dealer will set the bid and offer to recoup from liquidity traders what is lost to informed traders. In short, liquidity traders lose and informed traders win.

In this market there is a unique bid and offer. What makes this bid and offer unique is the anonymity of the investor. The market maker does not know whether the investor is uninformed or informed. If an investor could credibly identify himself as uninformed, that investor would receive a better price, and there would not be a single price. Thus, some investors would prefer nonanonymous markets in which they could obtain better prices. In a nonanonymous market the

transaction prices depend upon the identity of the buyers and sellers. There is no one unique fair price as there is in an anonymous market structured around a CLOB.

Indeed, this is exactly what has happened in the trading of stocks within mutual fund complexes and some other large institutional investors. For instance, firms like Vanguard or Barclay International routinely trade stocks from one fund or account to another in response to additions and redemptions.⁵ In the case of Vanguard, the trade price is the last trade. Since the last trade could be at the bid or the offer price, the average trade price will be near the midpoint of the bid and offer. Both funds will on average receive better prices than through the traditional markets. The trades are clearly liquidity motivated and should receive better prices than they would in the more general anonymous market.

Quite apart from reasons of anonymity, institutional investors often have orders that they feel will obtain better execution through markets other than the primary one. Institutions routinely use the upstairs market to trade large blocks of stock. The upstairs market maker may find another side for the trade to cross or act as principal and work the order later (Grossman 1992).

Innovation. Other methods of trading stock have developed to meet special needs. For nonurgent trades, institutions use crossing networks such as Posit or the Arizona Stock Exchange. Optimark has developed a very sophisticated call market in which institutions can post a “utility” function to dictate how their orders will be executed. Such innovation can lead to fragmentation, but it should be remembered that significant fragmentation would result only if the innovation is successful. If the innovation is truly successful, it may even drive other forms of trading out of business. After the transition process, there could be less fragmentation.

The Binding of Markets

Though it is natural for markets to fragment along the needs of different groups of investors, U.S. equity market centers have been bound together by the near simultaneous reporting of quote and trade information. Thus U.S. investors of both small and large means have reasonably good knowledge of what is happening in all market centers.

American investors now take for granted the consolidated reporting of information. Some major television stations routinely provide a sampling of last trade prices and volume of both

⁵ Neither the Consolidated Tape Association nor Nasdaq reports these trades.

NYSE-listed and Nasdaq stocks.⁶ The Web pages of many trading portals provide real-time quotes and last trade prices to investors.

Whether the U.S. investor has all the information that would be useful in evaluating market centers is an open question. There is no standard as to what information should be made available and at what time intervals, and markets differ widely in what is disclosed and when. For example, the U.S. corporate bond market is a fragmented dealers' market with virtually no trading information available to the public, even with delays. The London equity market allows a market maker to delay the reporting of large trades. Some markets, such as the U.S. futures market, do not report volume as it occurs.

The Structure of U.S. Markets Today

The NYSE and Nasdaq are the two primary listers of U.S. equities. As of December 31, 1999, Nasdaq listed 5,210 issues with a market value of \$5.2 trillion, and the NYSE listed 2,818 issues of common stock with a market value of \$12.2 trillion. Thus, the total market value of equities listed on both the NYSE and Nasdaq was \$17.4 trillion.

The Investors. According to the flow of funds, the total market value of all equities other than mutual funds held by U.S. investors totaled \$18.9 trillion as of December 31, 1999—somewhat more than the \$17.4 trillion reported by the two major market centers. Part of this difference may be due to direct holdings of foreign stocks by U.S. investors in excess of the direct holdings of U.S. stocks by foreigners. Another part of this difference may be due to the inclusion of nonlisted equities in the flow of funds.

Again according to the flow of funds, individuals held directly \$7.3 trillion or 39 percent of the total market value of equities held by U.S. investors.⁷ These direct holdings are highly concentrated, with millionaires holding 52 percent of all individually held stock (Goldstein and

⁶ The last trades reported are a filtered list of all trades. To display all trades would make the information unreadable. Still, this filtered list does provide a retail investor with sufficient information to judge whether a trade occurred at a reasonable price.

⁷ Table L.213 in the Federal Reserve statistical release "Flow of Funds Accounts of the United States," dated June 9, 2000, reports that households own \$7,829.4 billion of equities as of December 31, 1999 (<<http://www.federalreserve.gov/releases/z1/20000609/z1r-4.pdf>>). In the "Flow of Funds," households include nonprofits. In 1996 the household sector held \$4,642.1 billion, of which \$338.3 billion represents nonprofits. The year 1996 is the last year for which breakout is available. Thus, the equity holdings of nonprofits represent roughly 7 percent of the household sector. If they represent the same proportion today, the dollar amount owned by the household sector should be reduced by 7 percent to yield the number in the text.

Krutov 2000). In addition to these direct holdings, many individuals also have significant beneficial holdings of equities through mutual funds and pension funds.

Reliable estimates of individual trading are hard to come by. If one attributes all prints for NYSE-listed stocks of less than 2,100 shares to individuals, the trading of individuals would represent 7.5 percent of the consolidated share volume of NYSE-listed stocks as of August 2000.⁸ Interestingly, this share is somewhat less than the corresponding percentage of 9.9 percent in 1996. Since individuals' trades are generally small, individuals would account for a much greater proportion of the number of trades as distinct from volume.

Estimating retail volume on Nasdaq is difficult because institutions are more likely to split their orders into a number of smaller orders, and thus the number of small orders is not as good a measure of retail trading as it is for NYSE-listed stocks. However, Sanford C. Bernstein & Company has estimated that as of March 31, 2000, retail customers, including day traders, represented more than 60 percent of Nasdaq share volume (Bernstein Research 2000a). Consistent with Bernstein's estimate is the growth in trading volume for Nasdaq-listed stocks. The dollar volume of Nasdaq trading as of June 2000 was 122 percent greater than for the same month in 1999. In contrast, the share volume from the NYSE composite index increased only 34 percent over the same period. If most of this greater increase in Nasdaq volume is due to retail trading, particularly day traders, the Bernstein estimate might not be far from the mark. In any case, it is reasonable to conclude that retail trading, as a percentage of total share volume, is greater for Nasdaq-listed stocks than for NYSE-listed stocks.

In 1999, 78.7 million individuals owned stock directly or indirectly through mutual funds (ICI 1999, 13). Despite the press given to Internet trading, only 11 percent of traders used the Internet to buy or sell stocks in 1998. These individuals tended to be younger (with a median age of 41), more affluent (with a median investment in equities of \$127,600), and better educated than the typical investor (ICI 1999, 29). The percentage of investors using the Internet is undoubtedly greater today as the number of on-line brokerage accounts has doubled from 1998 through 1999.

In sum, institutional investors have investment discretion over a greater proportion of equity investments than individuals. In turn, rich individuals account for a disproportionate share of individually held stock. For NYSE stocks, institutions account for a much greater portion of

⁸ E-mail from George Sofianos, New York Stock Exchange, September 12, 2000.

the volume in comparison to individuals. In contrast, it now appears that individuals represent more than 50 percent of the trading of Nasdaq stocks.

In interpreting these retail trading numbers, it should be pointed out that only a limited number of retail customers trade actively. According to a recent survey, only 3 percent of individuals owning equities traded more than 12 times in 1998 (ICI 1999, 29). Thus, the high percentage of retail volume on Nasdaq is concentrated in a limited number of individuals.

In sum, investors in U.S. stocks are a disparate group that includes institutional investors, households with substantial assets, a large number of households with limited assets, day traders, less active traders, foreign investors, and so on. All of these investors represent different constituencies, and it should be expected, as it has occurred, that different ways of trading will evolve to satisfy their varying needs.

The Industry. The trading of stocks involves three primary functions: The first function is the gathering of trading orders, the second is the execution of these orders, and the third is the settlement of the trades. For the most part, these functions today reside in different organizations. The cost structures and the social externalities of these three functions differ, and each has different regulatory issues.

The firms that gather orders range from full-service brokers such as Merrill Lynch and Dean Witter to deep discount brokers such as LowTrades.com or Ameritrade. There are a large number of these firms. As of August 18, 2000, Yahoo listed 269 full-service firms, 148 discount firms, 27 day trading firms, and 103 Internet trading firms. Obviously there is some overlap among these lists, but the total is still a large number. That there are so many firms indicates that entry is easy and that there are not substantial economies of scale.

In contrast to the number of order-gathering firms, there are many fewer organizations that execute trades. NYSE-listed stocks are traded in various market centers, but the NYSE is the dominant market. In 1999 the NYSE captured 82.47 percent of the consolidated tape volume of NYSE-listed stock, the five regional exchanges captured 9.12 percent, and Nasdaq capture 8.41 percent. Since each NYSE-listed stock is assigned to a single specialist, there is currently very little fragmentation of trading in NYSE-listed stocks.

The trading of Nasdaq-listed stocks is spread over many more market centers. There are two principal types of market centers: market makers and electronic communications networks (ECNs). As of June 2000 there were an average of 12.7 market makers per listed stock; for the

top 1 percent of issues by dollar volume there were an average 52.9 market makers. Although these numbers seem large, there is a high degree of concentration of order flow in any stock in a limited number of market makers. Further, all of these market makers are linked together through electronic trading systems. SelectNet is a dealer communication system that allows dealers to trade among themselves. SOES is a small-order execution system that aggregates bids and offers of the Nasdaq market center, and any maker wishing to make a market in a stock has to participate in SOES for that stock.

A major barrier to a new market center is obtaining a critical mass of order flow. New markets face a formidable challenge in attracting the order flow needed to thrive: witness the Arizona Stock Exchange or Optimark. To displace an existing market requires that the new market center offer significantly better services than the established markets.

The settling of trades is even more centralized for any particular issue if the issue is registered either directly in the name of the owner or indirectly through an intermediary. In the United States, the Depository Trust & Clearing Corporation is the focal point for this transfer of ownership from one entity to another. In addition to handling the mechanics, the Depository Trust & Clearing Corporation guarantees the contra party risk. The problems that must be overcome in establishing a global settlement platform include (1) different settlement periods across nations interacting with differences in time zones, (2) the meshing of legal and regulatory processes among countries, and (3) minimizing the contra party risk in settling trades across nations. Of course, these problems are interrelated.

Regulation

In recent years the SEC has been extremely active in encouraging competition across market centers. These initiatives have had the most impact on Nasdaq, but they will ultimately have a major impact on the NYSE as well. There are three main initiatives: the Manning rules, order display rules, and alternative trading system (ATS) rules. Although each addresses different aspects of the market place, their total impact can only be understood through their interaction.

The Manning Rules. The Manning rules require that a market maker in Nasdaq stocks execute a customer's limit order before executing an order for its own account. Assume a market maker receives a limit order to sell 100 shares at \$20 and then receives a market order to buy 100

shares. If the current best offer for the stock were \$20, the market maker would have to execute the market order against the customer limit order before it could trade at \$20. Prior to the Manning rules, the market maker could have sold 100 shares to the market order from its own inventory at \$20, leaving the limit order unfilled.

The Manning rules thus assure a customer that if the customer has placed a limit order with a market maker and that market maker receives a market order that can be matched with the limit order, the limit order will be executed unless the market maker betters the price. If the market maker has two or more limit orders at the same price, they will be executed in time priority.

What the Manning rules do not provide is time priority at a given price across all market centers for Nasdaq stocks. Thus, an investor faces the possibility that a newly submitted limit order at the same price to another market maker will be exercised before his or her earlier-submitted limit order.

Under the Manning rules, the probability of an execution of a limit order depends upon the order flow received by the market maker. As the order flow to a market maker increases, the probability of a limit order being executed also increases. Prior to the Manning rules, there was no assurance of execution.

Order Display Rules. Under the Manning rules a market center had no obligation to display a limit order that improved the NBBO. Thus, if the NBBO were a bid of 20 and an offer of 20.50 and a market maker received a limit sell at 20.25, the market maker had no obligation to display the better offer. However, if the market maker received a market buy, the Manning rules required that the market maker exercise the order against the limit sell of 20.25 or improve the price.

The order display rules, which were published in 1996, require a market maker that receives a bid or offer that betters the NBBO to take one of three actions: (1) execute the limit order immediately against its own inventory, (2) display the better price as part of its own quote, or (3) send the limit order to another market maker who would then have the same three choices.

The order display rules also require that the quotes of ECNs be integrated into the NBBO. When the order display rules were made effective, the primary ECN was Instinet. Instinet is a market center used by market makers and institutions, and its best bid and offer were frequently better than those displayed by market makers.

The immediate effect of the order display rules was to narrow the spreads on Nasdaq stocks. This narrowing was undoubtedly due primarily to the narrower spreads on the ECNs.

Even if ECN quotes were not displayed, it is interesting to speculate about whether the spreads would still have narrowed merely from limit orders submitted by retail customers that would now be displayed.

The combined effect of the Manning rules and the order handling rules has probably increased the probability that retail and institutional limit orders that equal or improved the NBBO will be executed. As far as this author knows, there is no study in the public domain that has addressed the magnitude of this probability.

Alternative Trading System. The alternative trading system ruling made it easier to establish a new exchange. To date, no new exchange has been established under these rules, but both Island and Archipelago have applied to be exchanges. If one or both become new exchanges, this occurrence would create a direct competitive challenge to the NYSE. The NYSE has escaped many of the competitive changes that have impacted Nasdaq, but that will change with the existence of one or two new exchanges.

As an aside, both Island and Archipelago are currently ECNs, and it is useful to consider the difference between an ECN and an exchange. As mentioned above, ECNs are broker-dealers regulated by the National Association of Securities Dealers (NASD); they match public orders and do not act as principals. Both the ECN form of organization and an exchange have advantages and disadvantages. A major advantage of an ECN is that the NASD performs the regulatory function and frees the ECN from this activity. A major advantage of an exchange is that an exchange receives what is termed “tape revenue.” Such revenue comes from the sale of quote and last trade data to public vendors. These sales can represent from 20 to 40 percent of an exchange’s revenue. With an ECN the tape revenue goes to the NASD. Another advantage of an exchange is that it may have more latitude in setting its own rules.

To date, the quotes of all ECNs for NYSE-listed stocks, with the exception of Archipelago, have not been included in determining the NBBO for these stocks. If Island becomes an exchange, its quotes would become a determinate of the NBBO.

Competition, Fragmentation, and Time-Price Priority

A CLOB is the only way to maintain strict price-time priority across all limit orders. All orders would have to pass through this CLOB; otherwise, a crossed institutional order, for instance, might be exercised ahead of an already submitted limit order to that order’s

disadvantage. Such a CLOB would eliminate fragmentation, but it would also place a straightjacket on competition.

Competition is incompatible with a CLOB. Competition requires fragmentation. Price-time priority cannot be preserved in a fragmented market. Thus there is a trade-off among competition, fragmentation, and time-price priority.

Transparency across market centers helps to mitigate the effects of fragmentation. Last trade and quote reporting provide a great deal of transparency across U.S. market centers, but market centers are not fully transparent. For example, investors cannot compare market centers on their speed of execution or on the fill rate of limit orders. It is fully consistent with competition for regulators to assess the adequacy of current disclosure practices and then to require that more or less information be disseminated.

Strict price-time priority is inconsistent with competition. In a competitive world, an investor can expect only that his or her limit order will have a reasonable probability of being executed in a timely manner. What is reasonable and what is timely will vary from one investor to another and probably should not be regulated. However, the investor should have access to the information required to compare the quality of execution of limit orders across market centers.

For Nasdaq-listed stocks the Manning rules guarantee that a limit order submitted to any individual market center, a market maker or ECN, will be exercised according to strict price-time priority within that market center. Thus, the probability that a limit order submitted to a particular market center will be exercised in a timely fashion depends upon the magnitude of the order flow going to that market center. The magnitude of the order flow going to a particular market depends upon both the overall order flow in a particular stock and the concentration of that order flow among the top market makers.⁹

For NYSE-listed stocks the issue is more complicated. First, there is no attempt to preserve price-time priority across registered exchanges. Second, trading on the NYSE itself does not always conform to price-time priority across all trades, even though time priority within the limit order book itself is maintained. For example, a clean cross could take place at the price of a previously submitted limit order. Also, the trading rules of the NYSE sometimes permit a floor trader with a newly arrived order to trade in front of a previously submitted limit order.

That the NYSE sometimes allows trading ahead of previously submitted limit orders is not meant as a criticism of the NYSE. Trading practices on the NYSE are complex. Floor traders have certain advantages over those off the exchange, and one of these is that they can sometimes trade ahead of previously submitted limit orders. Giving certain advantages to floor traders may encourage the interaction of more orders on the floor of the NYSE, thereby decreasing fragmentation and increasing the probability that limit orders will be executed. Whether such externalities in enhancing liquidity exist is an empirical question.

In its proposed rules on order disclosures, the SEC has expressed a concern that payment for order flow may limit competition. It is a common practice for order-gathering firms to receive payment for sending their retail order flow to specific market centers regardless of their quotes. The market center paying for order flow agrees to execute the purchased order flow at the NBBO or send it to another market for execution. In the extreme case if all orders were preferenced, a market center would have no incentive to better the NBBO as such an action would attract no additional order flow.

Current trends in the market structure for trading stocks may mitigate any non-competitive aspects of paying for order flow. The Manning rules and the SEC order on the display of customer limit orders have gone a long way toward democratizing the trading process and removing the advantages that market makers have had, particularly in the trading of Nasdaq stocks. If the NBBO is a bid of 20 and an offer of $20\frac{1}{8}$, a retail investor who wishes to buy can either hit the offer and buy at $20\frac{1}{8}$ or put in a limit order to buy at $20\frac{1}{16}$. The market center receiving this limit order can execute it immediately against its own inventory, display the quote, or send it to another market center. If the market center retains the order without execution and then receives a market order to sell, the limit order would be exercised.

The probability that the limit order is exercised depends upon the proportion of order flow that the market center receives. Interestingly, under the Manning rules a market center that sends a limit order to another market center for execution is still obligated to execute that limit order if it receives a matching market order. In this case the market center that originally received the limit order must recall the limit order and execute it.

⁹ Additionally, if there is a demand for timely execution of limit orders, some market makers could respond by developing procedures to provide such execution as a function of trades occurring in other market centers. Madoff Investment Securities has stated that it has implemented such procedures.

Additionally, decimalization has begun on a pilot basis and will probably be extended to all equities in short order. Decimalization will certainly make it easier for the nonprofessional investor to understand prices, but the more important effect will be to reduce the minimum tick size from $1/16$ to one cent. When the tick size goes to one cent, the displayed spreads will probably narrow as some retail customers attempt to obtain better executions.¹⁰ With narrower spreads, market centers will pay less for each order, making payment for order flow less important in attracting order flow. Thus, at this time it is premature to prohibit payment for order flow. It may well be that the competitive rules that the SEC has already put into place and the growing use of screen-based trading by retail customers will make the issue of payment for order flow moot.

Global Trading

The U.S. equity markets are currently in a state of change. The SEC has unleashed new competitive forces and is very much concerned with market fragmentation, price-time priority, and competition among market centers. These concerns all deal with the domestic market. But the markets for U.S. equities are becoming increasingly global, and U.S. regulators need to focus their attention in this area.

Certainly within ten years, and probably within a much shorter period of time, the trading of the larger and more active stocks will be worldwide, at least at the retail level. Today, a Japanese investor can buy a U.S. stock on a U.S. market in U.S. dollars during U.S. trading hours. Tomorrow, the same Japanese investor will be able to buy a U.S. stock in yen during Japanese business hours. The physical location of the market center where the trade happens will not matter. The market center could be in Japan, the United States, or the Cayman Islands. Indeed, technology will make the market center's location, the currency, and the time of trade a matter of choice.

Today, a major hindrance to the development of a global market is the settlement and clearing mechanism. American Depository Receipts (ADRs) were developed to allow U.S. investors to buy foreign stocks on U.S. markets in U.S. dollars. A bank such as Bank of New York would buy shares of a foreign stock, hold them in trust, and issue an ADR representing a claim on a given number of shares of the stock. The ADR could be traded in U.S. dollars on U.S.

¹⁰As spreads narrow, the work of Goldstein and Kavajecz (2000) suggests that the depth of the quote may narrow.

markets. The bank would convert dividends in foreign currency to U.S. dollars and pay them. A holder of an ADR could obtain the underlying stock for a fee and with some delay. As a consequence the only real market for ADRs was in the United States.

Both the NYSE and Nasdaq are currently developing market structures to trade any stock in any nation in any currency. These new developments are in their infancy, but they point the way to the future. When Chrysler merged with Daimler Benz to form Daimler-Chrysler, the NYSE wanted to retain the primary listing and trading of the new company in New York, as did Frankfurt. An ADR was out of the question because it would recognize that Frankfurt was the primary market. The solution was to create a global share that could be traded in Germany and the United States in local currency and whose dividends were also paid in local currency. The fundamental hindrance to developing this global share was the settlement process. Germany has a $t + 2$ settlement period, and the United States has a $t + 3$ settlement period. There is also a difference in time zones. Thus, the settlement times of each side of a trade between a U.S. investor and a German investor are different. To bridge this gap, funds or stock need to be borrowed, increasing the cost of settlement and adding to contra-party risk. Additionally, the rules of the U.S. and German regulatory bodies had to be harmonized.

Nasdaq has formed an alliance with the Hong Kong Stock Exchange to trade some of the more active Nasdaq stocks in Hong Kong in Hong Kong dollars during Hong Kong business hours. That Hong Kong is twelve or thirteen hours ahead of New York depending upon the time of year makes settlement even more difficult to arrange. Nasdaq has also formed a joint venture to trade Japanese stocks and U.S. stocks in Japan. This venture will clear stocks traded between Japanese traders as if they were Japanese stocks. Only the net imbalance will have to be settled across time zones.

Once it becomes cheap and easy to settle trades across borders, there will be increased cross-border trading. But local markets will still have the critical order flow that will allow them to maintain their dominance. They will lose this dominance only if they do not respond to the needs of their investors or if domestic regulatory authorities impose excessively burdensome regulations. If so, other markets, perhaps in foreign countries, will develop.

The Future

The U.S. equity markets are in a state of transition. For the next several years, the impact of globalization should be minimal. The order handling rules, the enabling of alternative trading systems, and decimalization with the corresponding reduction in tick size have placed everything into play. In the next year the SEC may force market centers to provide data that will allow investors and order gatherers to make more informed decisions as to the quality of market making services.

It is likely that a new exchange will emerge. Without the historical baggage of existing exchanges, this new exchange may provide sufficiently attractive trading services to attract a critical mass of order flow. Alternatively, this new exchange may be the catalyst that causes existing market centers to react with better services, thereby preserving their market shares. In either case, investors will benefit.

In the longer term, national borders will no longer confine the trading of domestic securities to their home market. Today, the settlement process is a major barrier to the trading of U.S. equities worldwide. The technology is available to establish the order-gathering function or a market center anywhere in the world, but the cumbersome and costly settlement process across borders is the current barrier to locating these functions outside the United States. As market centers become more familiar with the settlement process across nations, the cost of settlement will go down. Further, just as banks in this country locate their credit card activities in states with favorable laws, order-gathering firms and market centers will be able to locate their activities in any country of choice. Likewise, U.S. investors will be able to enter orders through an order-gathering portal in any country.

The ramifications of global trading will be profound. With global trading, stocks of any country will be traded worldwide at low cost. It may be that the main market center for any stock will remain in the home country of that stock, but if the market center is not responsive to the needs of investors or if the rules and regulations in the home country become burdensome, the trading of the stocks of that country can and will move to another country.¹¹ The threat to move

¹¹When Sweden imposed a transfer tax in 1986—initially on stocks but subsequently expanded to options and futures—trading on the Stockholm Stock Exchange almost ceased as investors moved their trading activities to Wall Street and London. This tax was repealed on April 15, 1990, and trading returned to Sweden. This shift in markets occurred even without a global settlement system. In another example, in 1963 the United States imposed the interest equalization tax on foreign bonds purchased in the United States. Almost immediately the Euro-Bond

to another country will limit the ability of the SEC and other regulators to impose their wills. This threat would be mitigated if some investors preferred to keep their trading in this country in the belief that the U.S. regulatory framework provides enhanced protection against fraud, manipulation, and so on.

In a global market, U.S. regulators will face new and challenging issues. Examples include the following:

- How will the United States ensure the integrity of the settlement process across nations? How will it avoid a Metallgesellschaft-type default?
- Insider laws and their enforcement differ from one country to another. How will the United States enforce its own laws?
- If U.S. retail investors begin to trade non-U.S. equities in foreign markets, what steps should U.S. regulators take to protect such investors?
- Stockholder rights differ from one country to another. Will a company that discloses information in the United States according to U.S. rules be subject to litigation in another country that has different rules?
- What regulatory authority, if any, will protect a U.S. investor who buys a U.S. stock through an electronic portal located outside the United States?¹²
- The United States currently regulates margin through the lenders, rather than through the investors who utilize margin. Will individuals be able to bypass U.S. regulations on margin?
- In ensuring tax compliance, the United States relies heavily upon 1099 forms that report the dividends and sales receipts. In a global trading environment, how will the Internal Revenue Service maintain this reporting process?
- Although the equity markets in the United States are fragmented, the consolidated reporting of last trades and quotes provides a great deal of transparency. If a firm like Microsoft can be traded in any country, how can such transparency be maintained?

market was developed for trading such bonds to avoid this tax. Importantly, the market did not return to the United States when the tax was eliminated.

¹²The August 16, 2000, edition of the *Wall Street Journal* carried the story of a few European investors who were duped into overpaying for U.S. stock issued under Regulation S and were unable to recover any damages because of “regulatory limbo.”

As global trading becomes a reality, U.S. regulatory bodies will face new challenges. The concept of a national market system with time-price priority, or even simply price priority, will bend to the realities of a global market. It is imperative that U.S. regulatory bodies change their focuses from the regulation of a domestic equity market to the challenges of regulating a global market.

References

- Bagehot, Walter. 1972. The only game in town. *Financial Analyst Journal* 28, no. 1:80–84.
- Bernstein Research. 2000a. *Weekly notes*. New York: Sanford C. Bernstein & Co. March 31.
- . 2000b. *Weekly notes*. New York: Sanford C. Bernstein & Co. May 12.
- Glosten, Lawrence R., and Paul R. Milgrom. 1985. Bid, ask, and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics* 14, no. 1:71–100.
- Goldstein, Michael A., and Kenneth A. Kavajecz. 2000. Eights, sixteenths, and market depth: Changes in the tick size and liquidity provisions on the NYSE. *Journal of Financial Economics* 56, no. 1:125–49.
- Goldstein, Michael L., and Igor Krutov. 2000. *The future of money management*. New York: Sanford C. Bernstein & Co.
- Grossman, Sanford J. 1992. The informational role of upstairs and downstairs trading. *Journal of Business* 65, no. 4:509–28.
- Investment Company Institute and Securities Industry Association (ICI). 1999. *Equity ownership in America*. Washington, D.C.: Investment Company Institute and Securities Industry Association. <http://www.ici.org/pdf/rpt_equity_owners.pdf>.
- Securities Exchange Act of 1934 (SEA)*. 1934. *U.S. Code*. Title 15-Commerce and Trade. <www.law.uc.edu/CCL/34Act>.