January 25, 2005

Mr. Jonathan G. Katz
Secretary
Securities and Exchange Commission
450 Fifth Street, NW
Washington, DC 20549-0609

Re: File No. S7-10-04 – Regulation NMS

Dear Mr. Katz:

On December 16, 2004, the Commission has reproposed amendments to Regulation NMS (Proposing Release No. 34-50870, hereinafter "Release")[1] with the objective of modernizing and strengthening the regulatory structure of the U.S. equity markets. This comment focuses on the reproposed "Trade-Through Rule," which would require trading centers to establish, maintain, and enforce written policies and procedures designed to prevent the execution of trades at prices inferior to quotations displayed by other trading centers which are immediately and automatically accessible and are therefore protected by the Rule. The Commission considered two alternative versions of the Trade Through Rule: The "Market BBO Alternative," which would protect the best bids and offers of the nine self-regulatory organizations and Nasdaq (hereinafter, "BBO Alternative"), and a "Voluntary Depth Alternative" which would protect, in addition to all of the quotations protected under the BBO Alternative, all depth-of-the-book quotations at prices below the best bid or above the best offer that were designated as protected bids or offers (hereinafter, "DOB Alternative").

In this comment, we analyze and compare the two alternatives. Our analysis focuses on the listed equity market, although the principles we use have a broader applicability. Our work benefited from both partial financial support and access to information provided by the New York Stock Exchange ("NYSE"). The analysis below reflects our own independent views and conclusions, not those of the NYSE.

Our analysis follows.

Sincerely,

Yakov Amihud
New York University

Haim Mendelson
Stanford University

---

[1] Page numbers cited below are based on the original December 16, 2004 Release, not the *Federal Register* version.

Comment on Regulation NMS – "Trade Through" Rule
File No. S7-10-04

By

Yakov Amihud and Haim Mendelson

Summary of Conclusions

- Setting aside the costs of implementation, the BBO alternative is an improvement over the *status quo* whereas the DOB Alternative is worse than the *status quo.* The DOB Alternative mandates excessively tight coupling among market centers, which is likely to stifle innovation and commoditize execution services in the U.S. equity markets.

- Setting aside its market structure effects, the risks of implementing the DOB Alternative are substantial. If developed, it represents formidable operational challenges. The DOB alternative will substantially increase system capacity requirements and increase response times.

- We conclude that the DOB Alternative is undesirable from either perspective and is by far dominated by the BBO Alternative.

Introduction: Basic Principles

The Trade Through Rule is the result of a balancing act between two key principles that have guided the evolution of the U.S. equity markets over the years: The order/market interaction principle, which is intended to reduce market fragmentation and facilitate best execution for investors' orders (especially for retail investors, who do not possess sufficient resources and expertise to professionally evaluate the quality of the executions they receive); and the principle of market center competition, which promotes innovation, advances the securities markets through successive generations of technology evolution, improves execution quality and reduces execution costs to investors. Balancing the two principles has been a hallmark of U.S. securities regulation over the years, and that balance has served the U.S. markets well. The result was increased liquidity and lower transaction costs that reduced the cost of capital in the U.S. equity markets.[2]

[2] For an analysis of the effects of liquidity on the cost of capital, see, e.g., Amihud and Mendelson (1986a, 1986b). All references are listed at the end of this document.

The Utility Model vs. Intermarket Competition

A singular focus on the first principle, i.e., on order/market interaction, is consistent with the view that the U.S. equity markets should essentially function as a utility that provides efficient, undifferentiated services to investors. This model views the execution services provided by market centers as a commodity with no role for innovation. Because much of the cost of operating a market center is fixed, and because securities markets are characterized by "network effects" or positive liquidity externalities (Mendelson (1985)), the implication is that securities markets should be consolidated into a single, low-cost "utility." Under this view, all orders interact with one another, typically through the implementation of a Consolidated Limit Order Book ("CLOB") that concentrates all limit orders. Each incoming order is checked against the CLOB; market orders are executed against the best limit orders on the contra side of the book (following price-priority, followed by secondary priority rules), and limit orders are executed to the extent they are marketable or they are stored on the CLOB.

An attractive feature of this utility model is that it avoids the costs of market fragmentation. A market is fragmented when orders in the same financial instrument are decomposed into distinct subsets that do not fully interact, such as trading the same stock in two independent markets. The result of market fragmentation is that potential mutually-beneficial trades are missed, and other things being equal, the quality of execution is inferior to that provided by a market where all orders interact. Market fragmentation is costly, as it reduces liquidity and increases overall trading costs, hampers price discovery and reduces the incentive to provide information to the market (See, e.g., Cohen, Maier, Schwartz and Whitcomb (1982), Cohen, Conroy and Maier (1985), Mendelson (1987), Amihud, Lauterbach and Mendelson (2003)). Further, fragmentation can undermine best execution and market integrity when trade prices can be worse than the best available bids and offers or when a class of traders receives inferior treatment. Fragmentation can also lead to inefficient price discovery (Mendelson (1982, 1985, 1987), Amihud, Lauterbach and Mendelson (2003)).

While the utility model may seem attractive, it assumes a static market design and ignores the beneficial effects of intermarket competition in fostering innovation and reducing execution costs. Under the utility model, there are minimal incentives to innovate, as all trading is concentrated in a central utility. It assumes "one size fit all" preferences and has no role for alternative trading mechanisms that enable market participants to trade off different aspects of execution quality such as price improvement opportunities, market impact costs, size/depth and speed of execution. Intermarket competition induces each market center to improve its trading systems in order to increase or protect its market share. This effect is particularly powerful in the U.S. equity markets, where the differences in execution costs between competing market centers are razor-thin.[3] Over the past thirty years, competition drove markets around the world to

[3] For example, for most S&P 500 listed stocks, the difference between the lowest- and second-lowest effective half spread across the top three market centers that trade the

introduce automation and modern trading practices that facilitated investors' access to markets, improved liquidity and reduced execution costs. As is usually the case under competition, it also reduced rents due to market power and further improved liquidity.

The strength of the U.S. equity markets derives from their successful reliance on intermarket competition. More broadly, competition is the hallmark of the industrial organization of the U.S. economy, which sets it apart from centrally-planned economies. The U.S. willingly tolerated seemingly wasteful competition with the view that in the long run, the benefits of competition outweigh its costs.

Historically, the U.S. has rejected the utility model, favoring a regulatory approach that was based on intermarket competition. This approach has served the U.S. equity markets well, resulting in tremendous innovation that was accompanied by a fast decline in execution costs. Chordia, Roll and Subrahmanyam (2001) and Van Ness, Van Ness and Warr (2004) found that both the quoted and the effective spreads on NYSE listed stocks and on Nasdaq, respectively, fell dramatically over the 1988-1998 (for listed stock) and 1993-2002 (for Nasdaq) periods. In both cases, trading volumes increased substantially over the corresponding sample periods. According to the Plexus Group, the total cost of institutional trading in the U.S. declined by about 40% between 2000 and 2004 (Santoli (2005)). Also, data from Elkins/McSherry show that from 1996 to the 12 months ended June 2004, institutional market impact costs for listed stock in the U.S. declined by 63%, from 20.8 basis points in 1996 to 7.7 basis points by mid-1004; the corresponding decline on Nasdaq was 74% (from 50.7 to 13.0 basis points).

Much of the innovation taking place in the U.S. equity markets is due to intermarket competition. Santoli (2005) summarizes how competition and technological innovation drove these improvements: "It is cheaper, faster and more efficient to trade today than it's ever been... This is thanks not to decision makers in Washington, Boston or New York, but rather to long-developing shifts in technology and vigorous competition that began eroding the old ways of trading equities…" Indeed, much of the success of the U.S. securities markets has been achieved through the incentives created by multiple, competing market centers. The combination of low cost, improved speed and a greater variety of execution options could not have been achieved under the utility model.

Loose vs. Tight Coupling and Innovation

While unconstrained competition encourages innovation, there is often a strong case for imposing *some* constraints that link together different competitive offerings because of externalities, a potential reduction in transaction costs, or the achievement of desirable social goals. Given the tradeoff between the benefits of uniformity and scale and the benefits of competition and innovation, the U.S. regulatory environment has chosen the golden path of constrained competition, i.e., competition constrained by regulation that is intended to address issues that are not fully resolved by competition (for example, in the

---

stock is less that half a cent per share. This is based on a May 2003 sample; see Amihud and Mendelson (2003).

case of the securities markets, fragmentation). Implementing this view in the securities markets, the regulatory requirements create a degree of *loose coupling* among market participants, which is designed to achieve some of the benefits of coupling and coordination while preserving a high degree of independence and flexibility, so firms or market centers can innovate on their own in spite of being coupled or constrained.

A simple example is the publication of last sale information, which improves transparency and price discovery and provides price signals that can be used by both market participants and the entire economy for timely decision-making.[4] While the publication of transaction prices may make some traders worse off, and they may not have chosen to publish them voluntarily, its social benefits exceed the social costs and, importantly, the publication requirement does not interfere with the actual trading mechanism that had generated these prices—allowing market centers to innovate while adhering to the requirement.

Similar issues arise in the formation of standards, which on the one hand create constraints on product or service offerings and on the other hand can facilitate interconnection or coordination.[5] For example, the Internet Protocol enables multiple networks to interconnect and exchange data without imposing constraints on the types of data (and information) which are communicated over the network. The result is that each network node is free to innovate, while all nodes share the benefits of a universally interconnected network. When the Internet was originally conceived, its designers had in mind a telephony-style network like SNA and TYMNET. Had the Internet been optimized for its original applications, it would never have enabled the experimentation that led to protocols that could support the World Wide Web or the flexible interconnection that led many independent Internet Service Providers to emerge. Keeping the standards requirements to a minimum while preserving the ability to innovate on the network nodes quickly, at low cost and with minimal coordination has had very substantial benefits—loose coupling was better than either tight or no coupling.[6]

In both examples (the publication of last sale information and the Internet), market participants are loosely coupled, which achieves the benefits of *some* coupling (connectivity and interoperability in the first example, and price transparency in the second) while preserving a high degree of independence and flexibility, so firms or market centers can innovate on their own with minimal or no coordination, in spite of being coupled. Creating tighter coupling in an attempt to optimize a short-term objective is short-sighted, as it ignores the costs of complexity and interdependence, and it

---

[4] See Garbade and Silber (1994) for a historical overview.
[5] There is a vast literature on the economics of standards, lock-in effects and innovation; examples include David (1985), Besen and Farrell (1994), and Farrell and Shapiro (1992). Standards create benefits, but they also impose constraints that can hinder the ability of both new and existing players to innovate. This may result in an entire industry being locked into a solution which is suboptimal, as it was not the outcome that emerged from a fully competitive market test.
[6] See Saltzer, Reed and Clark (1981), Reed, Saltzer and Clark (1998).

jeopardizes long-term innovation (see, e.g., Iansiti and Levien (1994) for an overview). In a loosely-coupled system, each participant has the flexibility to innovate on its own without requiring agreement by the other participants, as long as the minimal (loose-coupling) standards are met. A tightly-coupled system creates substantial interdependencies that impede innovation and change. Innovation in a tightly-coupled system requires agreement by multiple parties, each of which can block progress. And, when the parties' interests are not aligned, for example when they compete, stalemate is the rule rather than the exception.

Effects of the DOB Alternative

Scope of Protected Orders

Under the DOB alternative, in addition to protecting the best bid and offer, all DOB quotations that markets publish to the consolidated quotation stream would receive trade-through protection. While the designation of a limit order as being protected is voluntary, it is predictable that in equilibrium, virtually all limit orders that are exposed in any market center will require DOB protection across all market centers. Obviously, once a market participant has decided to place a limit order in the market, she will maximize the probability of execution by having it protected in all market centers. Hence, her incentive is to have her limit order receive DOB protection.

If the decision on whether or not to designate limit orders as protected is made at the discretion of the individual market participant, she would thus opt for DOB protection of her limit order across all market centers. If the decision on whether or not to designate limit orders as protected is made by the market center itself (i.e., as part of its trading rules), market participants will choose to route their limit orders to the market centers that provide the desired protection for their limit orders. The result will be that virtually all limit orders will require DOB protection. Ultimately, the market will resemble a CLOB made up of all displayed limit orders from the different market centers that trade the stock.[7]

Sophisticated vs. Retail Traders

It is well known and widely-recognized that markets differ in various dimensions of execution quality – in particular, speed, trading rules (e.g., types of orders available) and reliability of execution (firmness of quotes). In addition, markets may differ in access, information provided (full book, top of the book, hidden orders, etc.), the interface they provide to investors and the systems that access the market centers. Open and flexible

[7] As a matter of implementation, the CLOB may actually be implemented via the publication of quotes to the different market centers (see discussion of implementation issues, below). Effectively, however, the system will resemble a CLOB or a "virtual CLOB."

interfaces increase transparency, facilitate the interaction between investors and the market center and enable investors to design trading strategies that they can carry out in the market with the most suitable interface. These considerations are particularly important for professional and institutional investors, who trade large quantities of stock, incur substantial market impact costs, engage in sophisticated trading strategies and often interact with the markets more directly. And, while a small, non-professional retail investor will most likely benefit from execution at the best price, institutional investors will not necessarily benefit from mandated trading at the best price. Sophisticated large investors have the ability to obtain execution at the best price on their own, while retaining the flexibility to consider additional tradeoffs that cannot be captured by a "one size fit all" rule.

The design of market structure rules needs to balance the benefits of rules designed to improve overall market performance (e.g., by improving the execution quality provided to retail investors, or creating greater incentives to place limit orders) against the costs (e.g., the costs borne by institutional investors who are constrained by the rules). While marginally constraining sophisticated investors may be justified if it improves overall market liquidity, as the constraints become increasingly stringent, the costs start exceeding the benefits.

The question is, then, what is the right balance between the costs and benefits of each set of market regulation rules. The decision should take into account market realities. First, most trading volume in the U.S. equity markets is due to sophisticated and institutional traders. Hence, excessively restricting their ability to make choices beyond a limited threshold may destroy more value than it creates. Second, sophisticated and institutional investors can route their orders off-shore, or to private or minimally-regulated trading venues, taking liquidity out of the public U.S. equity markets and thereby undermining overall market performance. We observe that this has not happened under the existing "trade through" rule in the public listed markets. As reproposed by Regulation NMS, the BBO Alternative will make the public listed markets even more attractive for sophisticated traders who seek fast electronic execution. Thus, it stands to reason that the BBO Alternative, as reproposed, will keep sophisticated and institutional traders in the public market. In essence, the BBO Alternative becomes a relatively minor inconvenience for institutional investors (since it will affect only a small part of a large order) while providing great benefits to retail investors.

Importantly, the protection of small investors under the BBO Alternative supports the viability of the market. It is well known that the existence of uninformed traders (sometimes called "noise" traders in the literature) is essential to make trading feasible in a market where there are also informed traders. It is through the trading of the latter group of investors that the price discovery process takes place, in the sense that the information available to informed traders is incorporated into stock prices. But trading will not take place unless there is a sufficient number of uninformed traders as well. Therefore, the market must be designed so as to accommodate the participation of these traders. The BBO Alternative can be viewed as a rule that protects and encourages the participation of small, unsophisticated traders in the market, which increases liquidity and hence benefits all market participants.

In contrast, the DOB Alternative imposes substantial constraints without providing commensurate benefits. Effectively, it commoditizes the quality of execution, turning it into a uniform product. As discussed above, this will adversely affect institutional traders, who benefit today from differentiation and choice, by limiting their choices. Furthermore, as argued above, if execution services are turned into a commodity, market centers will have no incentive to invest in improving their quality of execution on any dimension except price (e.g., new trading mechanisms, software, user interface etc.), since this will not generate a sufficient order flow to justify the investment.

Effect on Incentives to Innovate

The DOB Alternative will discourage innovation by preventing the market that invested in a liquidity-enhancing innovation from reaping its benefits. This will diminish the incentives for markets to compete by offering innovative trading procedures and will thus harm the development of the U.S. equity markets. Consider a sophisticated trader who wants to get the best price for her order. Because the DOB Alternative commoditizes the quality of execution, the trader will be practically indifferent as to what market center she submits her order to, knowing that even if she sends her order to the least liquid market center, it will obtain the best price available system-wide. If one of the market centers managed to innovate in a way that improves liquidity, this may not give it a competitive advantage, and the benefits of the improvement will be reaped by all market centers. As a result, a market center that contemplates investing in innovation or liquidity improvements will end up not making the investment. In the long run, the result will be to greatly dilute intermarket competition and diminish the incentives of all market centers to innovate and improve their quality.

Problems in the Introduction of New Trading Mechanisms

Imposing uniform standards across markets, as under the DOB Alternative, can freeze innovation by assuming that future applications will remain similar to present day applications. When an innovation represents a radical departure from the way business is done at present, it is impossible to foresee its future path. When standards are optimized for *current* applications, it is difficult to get around them and develop the applications of the future. As discussed above, the World Wide Web would probably have never come to life if the Internet Protocol had been designed for the telephony applications of its time. In the case of trading systems, a key dimension of innovation is the trading mechanism, namely: How orders are converted into trades. A simple example will illustrate how the DOB Alternative can constrain the development of new trading mechanisms by assuming the trading mechanisms which are prevalent at present.

As an illustration, consider the way execution prices are determined in the marketplace and the problems of introducing a different approach under the DOB Alternative. In principle, there are two basic methods for determining the execution price of a large order whose size exceeds the size of the contra bid or offer (so the order may needs to start sweeping the limit order book): Discriminatory pricing and uniform pricing. Under discriminatory pricing, each limit order (bid or offer) is executed at the limit price of that order. Under uniform pricing, all limit orders at a price better than or equal to the price of the last executed limit order receive the same price.

Discriminatory pricing is usually the rule on Nasdaq, while uniform pricing is used at the opening transaction of many stock exchanges, including the NYSE. A number of foreign stock exchanges implement uniform pricing both at the opening and during the day.[8] These two pricing methods are consistent with two leading auction methods, namely the discriminatory price auction and the uniform price auction, and there are well-known equivalence theorems showing under what conditions the two give rise to the same revenues (e.g., Harris and Raviv (1981), Milgrom and Weber (1982)). In practice, there is no unanimity regarding which method is better, although the U.S. Treasury, which used to auction Treasury securities using a discriminatory price auction, switched in the nineties to the uniform price auction and its studies suggest this improved performance at the auctions (Malvey, Archibald and Flynn (1995), Malvey and Archibald (1998)).[9] Whereas discriminatory pricing is widely used, uniform pricing may potentially improve market performance.

Under discriminatory pricing, if the size of a market or marketable limit order being executed exceeds the size of the limit order it hits, the incoming order is executed against the limit order at its limit price, and the remaining size moves to the next-best limit order on the book, and so on. That is, a large-size market (or marketable limit) order is executed in steps, with each part being executed at a different limit price. To illustrate, consider a given limit order book which is depicted in columns (a) and (b) on the table below. A market order to sell 9,000 shares arrives. Under the discriminatory pricing method, it will be executed as depicted in column (c), that is, 1,000 shares against limit order 1, 2,000 shares against limit order 4, 3,000 shares against order 3 and the remaining 3,000 shares against limit order 4. The total proceeds to the seller will be $64,000.

[8] See, e.g., Amihud and Mendelson (1987, 1991). The NYSE uses modified forms of uniform pricing in its block trading procedure and in its Hybrid Market proposal.
[9] There are many other studies on this issue, which need not be listed here.

Order bin	Price to sell	Size	Discriminatory pricing	Uniform pricing
	(a)	(b)	(c)	(d)
1	$9	1,000	1,000*$9	
2	$8	2,000	2,000*$8	
3	$7	3,000	3,000*$7	
4	$6	4,000	3,000*$6	9,000*$6
5	$5	4,000		
6	$4	4,000		
Total proceeds to seller			$64,000	$54,000

Now, assume the same limit orders and change the method of execution to uniform pricing: The sell market order of 9,000 shares is executed against all available limit orders in the order of their prices, and the transaction price is set according to the limit price at which the *last share* is executed. By this method, the entire sell market order is executed at $6 and the total proceeds to the seller are lower, only $54,000.

A naïve analysis of the two methods of execution may conclude that discriminatory pricing is superior for the seller since it generates greater seller proceeds. However, this is not necessarily the case, since it ignores the equilibrium behavior of traders in this market. Under the uniform pricing method, traders who place bids in bins 1, 2 and 3 know that they are likely to obtain *price improvement* in case a large order is executed. That is, instead of buying the stock at the price that they have quoted, they may buy it at a lower price than their limit price. In general, traders who place limit orders are apprehensive of their orders being executed against traders with superior information. In this example, the limit order trader in bin 3 who has placed a limit order at $7, may have valued the stock at $8 or more; but because of the risk of being picked off by a trader who knows that it is worth only $6 or less, he has quoted a price of $7. But under the uniform pricing method, the trader at bin 3 knows that his order may be executed at a better price than he has quoted. He therefore may be less apprehensive, and may increase the size of his limit order or raise his quoted price. Consequently, under uniform pricing the limit order book will be different from the book under discriminatory pricing: Traders who place buy limit order will quote a larger size at each price, or quote a higher price than they would under the discriminatory pricing method.

Now suppose a market adopts the uniform pricing method. It could be argued that traders can undo that and have their orders executed as they would under discriminatory pricing. The seller in the above example always has the freedom to break up his order of 9,000 shares in an attempt to achieve the discriminatory pricing result—even if the exchange implements a uniform pricing mechanism. He can send one market order of 1,000 shares, which will be executed at $9, followed immediately by another market order for 2,000 shares, which will be executed at $8, and so on, hoping to realize total proceeds of $64,000. However, this strategy is risky, especially in a fast-moving market. After the seller had sent the first order for 1,000 shares that was executed at $9 (proceeds of

$9,000), another trader may send a sell market order of 9,000 shares that is executed at the uniform price of $6. The seller in the example will then have to send 4,000 shares to be executed at $5 and another 4,000 shares to be executed at $4. His total proceeds are then $45,000. Had he immediately executed his entire order of 9,000 shares at the uniform price of $6, he would have realized $54,000. The potential higher cost to the seller from breaking up his order may thus inhibit him from simulating a discriminatory price sale when the market employs the uniform pricing method.

Uniform pricing is a perfectly reasonable way to price securities and it may form the basis for trading mechanism innovation in a securities market (as was the case in the U.S. Treasury Securities auction). However, such an innovation is entirely undermined by the DOB Alternative, which locks the market into a discriminatory pricing mechanism—and locks out new trading mechanisms that are based on uniform pricing. To see why, note that if the market center in our example wants to execute the order for 9,000 shares at a uniform price of $6, it must first clear the protected books of all other market centers at prices above $6, even though the executing market center itself has on its own book orders at better prices. This implies that a market center that chooses to adopt a uniform pricing scheme will lose its limit orders to other market centers that apply discriminatory pricing. In effect, *the DOB Alternative "freezes" discriminatory pricing as the trading mechanism of choice*. A more benevolent regulatory regime would allow market centers and market participants to make the choice as to the preferred pricing method. One would expect that it is intermarket competition that should determine which trading mechanism is better, and at the end the market centers with the better trading mechanisms will prevail. Or, multiple trading mechanisms will survive, each catering to a different clientele of traders who will choose what trading mechanism is best for them. Under the DOB Alternative, however, regulators effectively pick winners and losers, stifling innovation.

The above example illustrates a more general point. Innovative trading mechanisms that base their pricing on multiple limit orders are incompatible with the DOB Alternative, which considers one order at a time and insists on protection on an order-by-order basis. This applies to straddle-like orders, contingent orders to sell one security and buy another based on the prices of *both* securities, index-contingent orders etc. An entire set of innovative trading mechanisms would be eliminated if the DOB Alternative is adopted.[10]

Summary

The DOB Alternative seeks to obtain short-term gains while sacrificing long-term market development via innovation. And, even if all markets and trading systems were frozen in their current mode of operation, it is unclear that the DOB Alternative would be beneficial.

---

[10] Investors can still emulate some contingencies via individual orders which are enabled depending on multiple conditions—*after* these conditions were satisfied. But the whole point of trading mechanism innovation is to have competing market centers develop and support new types of orders rather than have investors create fragile workarounds. Also, note that uniform pricing requires a modification under the BBO Alternative, but that modification is minor (uniform pricing can be applied after "taking out" the BBO).

From an investor's viewpoint, while the DOB Alternative is intended to improve the quality of execution, it does not accomplish this objective. The executions that result from the DOB Alternative can be home-made by investors. They can design their own software that breaks a large trade and send the pieces directly to the various markets according to their displayed prices, as many smart routers are doing today. And yet, the DOB Alternative imposes costs, preventing investors from directing their orders for execution to the markets of their choice that provide them with other dimensions of execution quality beyond price. As such, the DOB Alternative is welfare-reducing: It adds nothing yet it constrains choice.

The DOB Alternative is harmful for long-term market development. It makes it difficult or even impossible to develop innovative trading methods that will improve the quality of execution. Because it reduces investors' incentives to search for the best market, it also reduces the incentive of market centers to develop a trading environment that is attractive to traders. An innovative market will gain very little by doing so. In sum, the DOB Alternative will hurt the development of the U.S. equity markets and the continued reduction in trading costs, driven by innovation and competition.

The BBO Alternative

In our view, the BBO alternative is the most effective way to link prices across different market centers so as the achieve the benefits envisioned by Congress from intermarket linkages, namely: "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 best execution of such orders" (Exchange Act Section 11A(a)(1)(D)).

The most reasonable balance between these elements—efficiency and competition on the one hand and obtaining best execution of investors' orders on the other—is achieved both effectively and efficiently by protecting just the best price, just the displayed size and only quotes that are available for automatic execution. This combination *protects small, unsophisticated investors*—those who need it most—while minimizing the scope of regulation and the associated constraints on market development. Importantly, by making the market more hospitable to small, uninformed investors, it increases liquidity. Further protection beyond the BBO of the interests of large, sophisticated investors is unnecessary as they can protect their interests on their own while it imposes constraints on market development.

Enforcement of the BBO Alternative strengthens and facilitates the implementation of broker-dealers' best execution obligations by effectively requiring the execution of small orders at a price which is no worse than the prevailing BBO. As the release recognizes, "Agency conflicts occur when brokers may have incentives to act otherwise than in the best interest of their customers. Customers, particularly retail investors, may have difficulty monitoring whether their individual orders miss the best displayed prices at the time they are executed. Given the large number of trades that fail to obtain the best displayed prices… the Commission is concerned that many of the investors that ultimately received the inferior price on these trades may not be aware that their orders did not, in fact, obtain the best price. The reproposed Trade-Through Rule would backstop a broker's duty of best execution by prohibiting the practice of executing orders at inferior prices, absent an applicable exception" (Release pp. 38-39). This analysis applies directly to retail investors, who tend to submit smaller orders of the type that will be protected by the BBO Alternative. These investors have neither the knowledge nor the resources to monitor the execution of their orders. And even if they had, the benefits of monitoring a small order may not justify the costs. For example, if an order for 100 shares is executed by a self-serving broker at a price that is 2 cents worse than the BBO, the savings from monitoring the agent is only $2.00 while the cost of time of doing so is greater.

On the other hand, institutional investors have the ability, knowledge and resources needed to monitor and optimize the execution of large orders, and they can easily trade off for themselves the different dimensions of execution quality as they apply to their orders. They hardly need the protection provided by the DOB Alternative, and many of them would prefer to be in complete control of how their orders are executed. In contrast, small investors are the ones who would benefit from the limited, and well-targeted, protection which is provided by the BBO Alternative. Protecting just the size available at the BBO in an automated fashion is likely to be sufficient for small retail investors while imposing a minimal burden on large institutional investors and market centers.

Second, the quotes that deserve the highest degree of protection are the best bid and offer, which form the basis for pricing most (small) transactions and ultimately constitute the price signals used by the economy at large.

Third, the implementation of the BBO Alternative is a natural continuation of existing trends and capabilities in the U.S. equity markets: The Trade-Through Rule obviously already exists for listed stock, and Regulation NMS will speed up execution and increase the net benefits of the rule. With the Order Handling Rules, SuperMontage and the ADF, Nasdaq has already moved towards the structure envisioned by the reproposed Rules, and implementing the BBO Alternative will further close the gap between the listed markets and Nasdaq.

Fourth, as discussed earlier, the protection of small, least-informed investors' interests is paramount for the viability of the market and for its liquidity. Kyle (1985) showed that the larger the trading volume of uninformed investors, the higher the liquidity of the market. Glosten and Milgrom (1985) showed that the greater the likelihood of trading with an uninformed investor, the narrower is the bid-ask spread, i.e., the greater is the liquidity of the market. It follows that regulations that induce more trading by uninformed investors are beneficial to market liquidity. By protecting small investors, the BBO Alternative accomplishes this objective.

Incentives to Innovate

While generating a number of benefits, the BBO Alternative does not inhibit the incentives to innovate. In the discussion of the implications of the DOB Alternative, we examined the effects of that Alternative on the incentives to innovate, and concluded that the DOB Alternative will hurt competition and impede innovation and progress. In contrast, the BBO Alternative, where only the top of the book is protected does not inhibit competition between markets in the development of innovative trading methods.

A trader contemplating a large trade would manage her trade in a way that is optimal for her. It would compare the different market centers and select the one where she can most likely obtain the best overall execution, even if part of her order will be executed at a worse price than in another market center. Or, the trader may choose to break the trade and execute it in multiple markets, thus home-making any benefit that could potentially accrue from the DOB Alternative. If the trading that would result from the DOB Alternative can be home-made by this trader, the rule is redundant and in fact it is welfare-reducing. This is because it robs the trader of the freedom to deviate from it, should her interests so dictate. In addition, the DOB Alternative causes a welfare loss by inhibiting competition. In contrast, under the BBO Alternative, markets will be responsive to traders' choice criteria. Each market will compete for order flow by striving to provide better liquidity and overall execution quality. Markets will adopt trading rules, priority rules, order types, technology etc. so as to attract a specific clientele for which they can provide the best execution. Since liquidity begets liquidity, this will attract greater order flow to the more liquid market, which in turn will attract more limit orders, since traders who place them will expect greater likelihood of these orders being executed quickly. The end result is that markets will have strong incentives to innovate, liquidity will be enhanced and traders will benefit from both.

This process helps a market become more liquid and consequently more attractive, and it incentivizes other markets to develop better trading systems that will turn them into winners. A case in point is the International Securities Exchange (ISE), which started trading stock options in 2000, at which time the Chicago Board of Options Exchange was the dominant market for equity options. As of the end of 2004, the ISE is the largest market for stock options. It succeeded by building an advanced and innovative electronic system and adopting rules that attracted trading away from other markets, induced market makers and traders to quote larger sizes, and enhanced liquidity. One of the rules that helped the ISE succeed gave priority to public customer orders: "Public Customer Orders on the Exchange shall have priority over Non-Customer Orders and market maker quotes at the same price in the same options series." (ISE Rule 713(c)). Whereas a common secondary priority rule is size, which pushes small public investors to the back of the execution queue, the ISE devised a trading rule that encourages the participation of public investors. Clearly, giving priority to public investors entails a cost, since it overrides other reasonable secondary priority rules, such as time and size. But by attracting smaller public investors, the ISE created a more liquid market.[11] Innovation (in both technology and trading rules) has paid off.

Conclusion

The BBO Alternative is most potent in protecting the interests of small, uninformed investors. This will induce their participation in the stock market and thus will make the market more liquid. At the same time, the equity markets will continue to be loosely-coupled and free to innovate as they have been in the past. Because the coupling is limited to the BBO, this Alternative provides the minimum degree of interference necessary to link the markets as required by the 1975 Exchange Act Amendments, while leaving them the flexibility to define new order types and develop new trading mechanisms with a minimal degree of interdependence with the competing market centers. This will foster intermarket competition, which was the strongest driving force behind the development of innovations in the U.S. capital market in the last decades.

Some Implementation Considerations

The discussion below addresses issues raised by the implementation of the DOB alternative. We first address system development issues, and then turn to briefly discuss a few performance and implementation issues.

Development Risks

The development of real-time systems is a complex, risky and time-consuming process. Even systems that do not have stringent performance requirements often fail: In a widely-cited study of 365 companies (Standish (1995)), the Standish Group found that only 16.2% of software projects studied were successful, while 31.1% of projects were canceled and 52.7% of projects were "challenged," i.e., they were completed over-budget, over the time estimate, or with fewer features and functionality than originally specified. While project success rates have improved over time, the Standish Group estimates of the $255 billion in 2004 software project spending in the United States, $55 billion will be wasted , made up of $38 billion from failed projects and $17 billion in cost overruns. Stringent performance requirements have long been known to increase development cost and risk (Boehm (1981)), increasing the probability of failure.

Unlike the BBO Alternative, which extends current practices and architectures that are already working in the marketplace, the development and implementation of the DOB

[11] This is analogous to the BBO Alternative for the Trade-Through Rule, which protects market integrity for small retail trades.
investors in the market, may have a small cost but its benefit in
attracting such investors makes it worthwhile to adopt.

Alternative would require putting together an entirely new architecture and the development and implementation of a new transaction processing paradigm. This creates significant development risks.

Some of the best known “risk factors” in systems development are presented by the DOB Alternative:

- The system is based on a new, untested paradigm, which stretches current performance standards.

- The system has a large number of interfaces, which increases interdependencies, complexity, and the probability of failure. This is closely-related to the discussion of “tight coupling,” where the issues discussed earlier translate into system development risks.

- The “customer” for the development project is multiple, competing market centers with conflicting objectives.

- The project focuses on a core activity which will affect the participants’ competitive positions. When multiple companies or organizations build a non-core system such as a back-office system, all of them stand to benefit from the success of system implementation, and that success is unlikely to affect the competitive position of any one company. Hence, there is a strong incentive to cooperate. When the system is a core system that will affect the participants’ competitive positions, systems design is closer to being a zero-sum game, which makes the design process part of their competitive strategies and increases the probability of failure.

Many of these characteristics (but not the stringent performance requirements) were shared by Taurus, the electronic share transfer and registration system that was intended to replace paper certificates and forms with computer-based processing for the London Stock Exchange in the early nineties. Taurus was a large, ambitious project with multiple industry participants who had to agree on messaging standards and operating procedures. It was canceled a few years after its initiation, when it became clear it was going nowhere; the development effort cost the London Stock Exchange about £80 million and the other industry participants about £400 million.

System Alternatives

The BBO and DOB Alternatives couple multiple exchanges in different ways. They require data, and potentially trading logic, to be shared across multiple market centers, which can be accomplished through messaging and/or a common database. Conceptually, the shared data may be **thin** or **thick** and it may be **centralized** or **decentralized.** Similarly, the trading logic may be **centralized** or **decentralized**. A change in any of these basic parameters is fundamental, as it changes the underlying system architecture, and as a result it requires substantial development effort, as discussed above.

"Thin" means that for each stock, the shared data consist of a limited number of data items such as the best bid and offer in each exchange (along with other data such as size and time), last trade information, etc. With thin shared data, the scale of the database and the associated systems needed to maintain the shared data is limited by well-defined bounds that can often be determined in advance. "Thick" means that the shared data (for each stock) may consist of a large, variable number of records, and the number may change across stocks and over time. For example, the DOB Alternative requires sharing the depth of the book at multiple price points across multiple market centers, which requires planning for a large number of records.[12]

"Centralized" and "decentralized" refer to where data are stored or trading logic is executed. Centralized means that a central system manages the shared data or trading logic. Decentralized trading logic means that trading is executed entirely at the systems of the individual market centers; decentralized data means that they share data through messaging.

When all orders are managed using a central, physical CLOB, the shared data is thick and (entirely) centralized, and the trading logic is centralized as well. A central CLOB is an efficient implementation of the utility model. It can guarantee execution since both the data and the logic are centralized.[13] When the CLOB is the only trading mechanism available, there is no substantive role for multiple exchanges — all of them would simply feed the CLOB and serve as "dumb pipes" between market participants and the CLOB. Trade execution becomes a commodity, the CLOB becomes a regulated monopoly, and there is only a minimal role for innovation.

The current system for trading listed stock is a system with decentralized trading logic and "thin" centralized data. The trading logic is decentralized because the Intermarket Trading System (ITS) is merely a communication system used to route messages across the different market centers, where the actual trading takes place. The shared data is "thin" since it is sufficient to maintain in a central database "top of the book" information

---

[12] Recognizing that the DOB Alternative may raise capacity issues, the Release (Footnote 119) suggests that "the SRO participants in the applicable market data Plans potentially could determine to disseminate only those DOB quotations that were within a certain number of price levels away from the NBBO." However, while the benefits of protecting DOB quotations accrue to the sending market center, much of the cost is incurred by the receiving market centers, which creates an incentive not to restrict DOB quotations in this manner. And, requiring market centers to limit DOB quotations to just those around the BBO defeats the purpose of the DOB Alternative and creates a new set of anomalies while still incurring the very high costs of design and implementation.

[13] A centralized CLOB may become a single point of system failure and congestion. However, there are well-known techniques for dealing with these problems, viz. backup facilities and load balancing.

for each participating market center, with DOB data being maintained only by the individual market centers. Because the shared data is "thin," it is possible to send it efficiently to all market centers: To each quotation disseminated by the Consolidated Quotation System (CQS), the system appends the NBBO for the corresponding security.

With the reproposed Regulation NMS, the BBO Alternative will preserve these fundamental attributes and architecture for listed stock. The key change is that the system will link automated execution systems with no manual intervention. This will keep the trading logic decentralized and the shared data "thinly" centralized. In other words, the exchanges will continue to compete within a loosely-coupled system, where each can independently innovate with limited reliance on the other exchanges.

Between January and November of 2004, approximately 2% of trades in the participating market centers were traded in the ITS (the monthly low was 1.9% in July 2004, and the monthly high was 2.8% in January 2004). ITS trading volumes hovered between 3% and 4% of total consolidated volume. About three quarters of commitments to trade sent to the ITS resulted in an execution (the monthly low was 72% in January 2004, and the monthly high was 77% in June 2004). As the reproposed regulation NMS will dramatically increase the probability of execution under the BBO Alternative, it is likely that intermarket trading volumes will increase, perhaps dramatically, requiring additional technology investments. However, this will improve performance without requiring a fundamental change.

This is not to say that the BBO Alternative will lead to no substantial change—rather, the major innovations are occurring within the exchanges rather than on the interfaces between them.[14] The NYSE has made its Hybrid Market proposal, a major trading mechanism innovation designed to operate within the context of trade-through protection for automated quotes that are available for immediate execution. Importantly, with tight coupling among the market centers, such innovation would not be possible without lengthy deliberations and negotiations with the NYSE competitors, which could add years to the proposed changes — or extinguish them altogether.

DOB Alternative

Unlike the BBO Alternative, the DOB Alternative would result in fundamental changes in the architecture of the markets for listed stock. Its most straightforward implementation is through a physical CLOB controlled by a monopolistic exchange or by a cartel. While we do not have sufficient information to estimate the costs of operating the DOB Alternative at this time, its implementation is likely to be expensive, inefficient and risky.

[14] This is similar to the way innovation takes place on the Internet: The locus of innovation is at the nodes rather than within the network.

Preserving a degree of competition among market centers requires the trading logic to be decentralized.[15] Because the shared data include DOB details, it is necessarily "thick." One can broadly distinguish between the case where the shared data is decentralized (i.e., the market centers transmit messages to one another) and the case where it is centralized (i.e., there is a central CLOB, although no executions take place at the CLOB itself).

Under the decentralized approach, there is no physical CLOB, and each market center sends its protected DOB orders to all other market centers. Each market center constructs its own local version of the CLOB by aggregating the protected DOB quotations received from all other market centers. Before a market center executes an order, it checks its local version of the CLOB so constructed to see if better prices are available in other market centers. If there are better prices, it sends Immediate Or Cancel (IOC) limit orders for execution to the market centers whose protected limit orders it had received. The executing market centers execute these IOC orders, report the executions or cancellations back to the sending market center, and send cancellations or updates for the limit orders that were executed to all market centers. The originating market, having received the cancellations, can update and execute or send new IOC orders again, and the process continues until the entire original order is executed.

If the shared data is centralized, a central database (CLOB) aggregates the protected DOB quotations across the nine market centers. Each market center sends each protected order, or each change in or cancellation of a protected order, to the CLOB. The CLOB stores the data in the form of a limit order book (by stock, price, market center, size, and possibly other information) and publishes it in compressed form at high frequency to all market centers.

Before a market center executes an order, it checks its local copy of the CLOB to see if better prices are available. If there are better prices on the CLOB, the market center sends Immediate Or Cancel (IOC) limit orders for execution to the market centers whose limit orders it found on the CLOB.[16] The executing market centers execute these IOC orders, report the executions or cancellations back to the sending market center, and send the changes to or cancellations of the limit orders that were executed to update the CLOB.

The implementation of this alternative depends on the way data is aggregated on the CLOB and compressed for transmission. The dissemination of its CLOB is somewhat similar to the NYSE OpenBook system, which is an electronic data feed that provides subscribers aggregate bid and offer information at each price point from the NYSE

[15] One may argue that the exchanges can compete even if their trading logic is consolidated, but there is no practical example of this, as the trading logic is a core element of what an exchange is.

[16] In an alternative implementation, the CLOB aggregates the limit orders for each stock at each price point. Under this implementation, the market center sends the orders to the CLOB, which parses them out into underlying limit orders which are then routed to the appropriate market centers for execution. This, however, calls for some distributed trading logic.

Display Book. The current implementation of OpenBook sends the aggregated content of the NYSE Display Book in the morning, and then sends out updates every 5 seconds. An upcoming version will send out an update once a second. The 5-second version of OpenBook transmits on average 300 messages per second (MPS), with a maximum of 500 MPS. The maximum message rate for the 1-second version of OpenBook will be about 1,200 MPS.

Under either approach, each market center will need to manage both its own limit order book and a local version of the CLOB; the difference is in how that local version is disseminated, constructed and maintained. Beyond a certain CLOB size, the centralized approach is likely to be more efficient, and may then be implemented in a way that is similar to the NYSE OpenBook.

Performance Issues

Trading systems are among the most complex transaction processing systems in existence, because they are high-volume and time-sensitive and because they require high reliability and fault-tolerance. Key performance measures for these systems include capacity, response time and the probability of execution. These performance measures are not independent, and in fact they are both substitutes and complements.

Most important is the tradeoff between capacity and response time. In general, the response time of a system depends on its capacity (the number of transactions it can process per second assuming no congestion, which in turn depends on the amount of processing required per transaction and the system's computing power), on system congestion, and on the network's latency, namely the amount of time needed for a message to travel across the network. Local processing, database access and update and network latency can each become bottlenecks that inflate response times. Because (for a given system capacity) response time increases faster than the increase in the load on the system, it is important to design systems with significant excess capacity for the key resources.[17] And, because system load changes over time, the system has to be designed for its peak load, which by far exceeds the average load.[18] Furthermore, greater variability or uncertainty about order inter-arrival times translates into larger capacity requirements for any given message volume.[19]

In addition, different system resources need to be balanced to optimize the use of capacity. Otherwise, a resource becomes a bottleneck that increases the response time regardless of the capacity of other resources. This is difficult to achieve when the load on the system is largely driven by other market centers, whose behavior is less predictable

---

[17] As the utilization of a congested system approaches 100%, its response time increases to infinity.
[18] In a typical day, the NYSE peak load is ½ to 2/3 higher than the average load, but on some days the peak load is twice the average load.
[19] Mendelson (1985) quantifies these relationships.

and uncontrollable. And, the less predictable the load on the system is, the greater the excess capacity required to achieve a desirable response time.

Another important relationship is between system response time and the probability of execution. Longer response times can reduce the probability of execution, because by the time an order arrives for an attempted execution, the limit order it attempted to hit may not be available any more — it may have been canceled or executed earlier against another order. This problem is particularly relevant given that the limit orders on some ECNs have extremely short duration. For example, Hasbrouck and Saar (2004) find that 25% of the limit orders submitted to the Island ECN were cancelled after two seconds, and above 40% were cancelled after 10 seconds.[20]

Also, the effective time to execution is longer than the response time per execution when executions are not guaranteed: If each attempted or actual execution takes a second and the probability of execution is 50%, it takes 2 seconds on average to actually execute the order.[21] Thus, short response times with a low probability of execution can translate into long effective times to ultimate execution.

Scale and Capacity Trends

On the NYSE, processing capacity is increasing at a much faster rate than the actual trading volume. For example, between the fourth quarter of 2000 and the fourth quarter of 2004, NYSE processing capacity increased by a factor of 8 (with the number of quotes and the number of orders processed by the system increasing by more than a factor of 8 over the same period) while the number of trades increased less than fourfold, and trading volume less than doubled. This was driven in part by decimalization and the associated decline in average trade size, and in part by the increase in computer-driven, algorithmic trades.

In the last quarter of 2004, the NYSE received about 15.6 million orders per day on average. Of those, 93%, or 14.6 million per day on average, were limit orders. The level of activity per trade is higher in the other market centers that trade NYSE-listed stock. To estimate the number of limit orders in these market centers, we used Rule 11Ac1-5 data for covered orders[22] over the 12-month period from November 1, 2003 through October 31, 2004 (more recent data were not available). During that period, for each

[20] The reproposed Regulation NMS would not extend protection to "flickering quotations" whose prices have been displayed for less than a second, but that will leave most short-duration quotations within the scope of the DOB Alternative.
[21] Under independent attempts with probability of successful execution of ½ for each, the expected overall time to execution is $(½)\bullet 1 + (½)^2\bullet 2 + (½)^3\bullet 3 + \ldots = 2$ seconds.
[22] Under Rule 11Ac1-5, a covered order is an order that was received by a market center during regular trading hours at a time when a consolidated best bid and offer is being disseminated, excluding orders with special handling instructions and orders that are more than 10 cents away from the BBO. Also, this assumes that the CLOB needs to maintain each limit order.

covered limit order on the NYSE, ArcaEx had 0.18 limit orders, INet had 0.19 limit orders, Island had 0.12 limit orders, and all market centers away from the NYSE combined had .55 limit orders on average.

With the above estimates, the size of the CLOB for NYSE-listed stock is about 1.5 times the size of the NYSE Display Book.[23] Since the number of orders is a key driver of capacity, this translates into a sizable increase in NYSE capacity requirements.[24] The required capacity increase on the other market centers that trade NYSE-listed stock will be much larger, since each will have to move from managing a limit order book sized at a small fraction of the NYSE Display Book to managing a CLOB sized at 1.5 times the NYSE Display Book. Even under the centralized shared data alternative, which shifts the cost of constructing the CLOB to a central processing entity that disseminates it to all market centers, each market center will need to query its local CLOB, to effectively merge it with its own limit order book, and to manage the message traffic sent to the CLOB as well as the orders sent to each of the other market centers. This translates into a large increase in processing requirements that needs to be multiplied by the number of market centers. In essence, each of the eight market centers will need to scale its capacity upwards towards the processing capacity of the NYSE—even if trading volumes remained the same.

Response Times and Probability of Execution

One of the tenets of Regulation NMS is assuring a standard of subsecond response time for electronically-executed orders. This standard will likely be defeated by a system that complies with the DOB Alternative.

First consider the timeliness of the local CLOBs. While a centralized CLOB may economize on capacity requirements, its dissemination results in additional delays.[25] At the current size of the NYSE Display Book, one-second updates represent a practical high-end, and this is a likely outcome for a larger shared CLOB. This translates into a one second difference between the central CLOB and its local copies. To this, one needs to add other components of delay due to incoming messages that update the CLOB, the time needed to update the CLOB, transmission delays and latency.

---

[23] This ratio is based on orders covered by Rule 11Ac1-5. However, more than 80% of limit orders on the NYSE were covered by Rule 11Ac1-5, so even if the ratio is different for orders that are not covered by the Rule, this would result in a small change whose direction is indeterminate.

[24] The actual capacity requirements depend on the actual workflow for these orders which is not known at this time.

[25] The corresponding delay under the decentralized alternative depends on the delays in messaging, as well as on the local CLOB processing times, which in turn depend on the capacities of the market centers' systems. A 1-second delay, similar to the centralized alternative, is plausible.

While it is difficult to estimate the actual response times, the data in each local CLOB will likely be a few seconds old, with order routing and execution taking a second or more (the time to execute an order on Direct+, the automated execution system of the NYSE, is approximately half a second, and this does not include the times of going to and from the NYSE; and the time to report an execution to the Consolidated Tape is about 1.5 seconds). This means that executions will not take less than a second and furthermore, traders cannot rely on the limit orders on the CLOB as being firm when they send orders to another market center for execution. Because a single order in a market center may translate into multiple orders sent to multiple market centers, the trader will have to wait for an acknowledgement from the slowest market center whose DOB orders were protected before being able to decide what to do next. The result is a further decline in the probability of execution which in turn slows down the effective time to ultimate execution.

These problems are exacerbated by the relationship between news, order placement strategies and order inter-arrival times (Engle and Russell (1998), DuFour and Engle (2000)). When news hits the market, orders tend to come *in sequence* on one side of the market, that is, we observe a sequence of buy orders or sell orders. With new information, there is a rush to trade on the news against the outstanding limit orders. A race develops both among multiple traders who want to quickly trade on the news, and between those traders—who want to hit standing limit orders on the book—and those who have placed the limit orders and continuously monitor the market to adjust or cancel them. During these periods, subsecond response time is crucial and order interarrival rates are short. But these are exactly the periods when the system tends to be congested, and response times will thus tend to be longer. The DOB Alternative requires orders to be executed against protected limit orders in other market centers, which increases the response time and reduces the probability of execution. During this period, there is a higher probability that a limit order was hit by another order or was canceled by the market participant who placed it, meaning that there is a higher probability that an order routed to another market will not be executed. The end result is that *congestion and delay are higher, and the probability of ultimate execution is lower, when the opportunity costs of time and of unexecuted trades are higher*, which magnifies both the response time problem and the associated losses. And, this problem is particularly serious exactly when fast trading is important, i.e., when traders rush to trade on new information.

In summary, while it is premature to estimate the actual capacity and response times under the DOB Alternative, the capacity requirements across the market centers are very substantial, and the need to share CLOB data results in a significant degradation in response times and the probability of execution.

Fundamentally, the most efficient way to implement the utility model is to build a central utility rather than replicate it within each market center to create a semblance of competition. With the centralized data, centralized trading logic implementation described earlier, a single processing facility houses and manages the CLOB as well as all trading. This obviates the need for excessive messaging, minimizes capacity requirements (because of scale economies) and delay (because intermarket communication becomes unnecessary, and because of the statistical properties of congestion) and maximizes the probability of execution. The logical outcome of the DOB Alternative, both from a market structure perspective and from an implementation perspective, is the utility outcome, namely a single, commoditized market that consolidates all trading. It should be rejected in favor of the BBO alternative, which preserves intermarket competition, fosters innovation and has served the listed equity markets well.

.

REFERENCES

Amihud Y. and H. Mendelson, Liquidity and Stock Returns. *Financial Analysts Journal*, Vol. 42, 1986, pp. 43-48.

Amihud Y. and H. Mendelson, Asset Pricing and the Bid-Ask Spread. *Journal of Financial Economics*, Vol. 17, 1986, pp. 223-249.

Amihud, Y. and H. Mendelson, Trading Mechanisms and Stock Returns: An Empirical Investigation. *Journal of Finance* Vol. 42, 1987, pp. 533-553.

Amihud, Y. and H. Mendelson, Volatility, Efficiency and Trading: Evidence from the Japanese Stock Market. *Journal of Finance*, Vol. 46, 1991, pp. 1765-1789.

Amihud Y. and H. Mendelson. Effects of a New York State Stock Transaction Tax. Working Paper, July 2003.

Amihud, Y., B. Lauterbach and H. Mendelson, The Value of Trading Consolidation: Evidence from the Exercise of Warrants. *Journal of Financial and Quantitative Analysis*, Vol. 38, 2003, pp. 829-846.

Besen, S. M. and J. Farrell, Choosing how to compete: Strategies and tactics in standardization. *Journal of Economic Perspectives*, Vol. 8, 1994, pp. 117-131.

Boehm, B.W., *Software Engineering Economics.* Prentice Hall, Upper Saddle River, NJ, 1981.

Chordia, T., R. Roll and A. Subrahmanyam. Market Liquidity and Trading Activity, *The Journal of Finance*, Vol. 56, 2001, pp. 501-530.

Cohen, K.J., R.M. Conroy and S.F. Maier, Order Flow and the Quality of the Market. in *Market Making and the Changing Structure of the Securities Industry*, Lexington Books, MA, 1985, pp. 93-112.

Cohen, K.J., S.F. Maier, R.A. Schwartz and D.K. Whitcomb, An Analysis of the Economic Justification for Consolidation in a Secondary Security Market. *Journal of Banking and Finance,* Vol. 6, 1982, pp. 117-136.

Cusumano, M. A., Y. Mylonadis, and R. S. Rosenbloom, Strategic Maneuvering and Mass-Market Dynamics: The Triumph of VHS over Beta. *Business History Review*, Vol. 66, 1002, pp. 51-94.

David, P. A., Clio and the Economics of QWERTY. *The American Economic Review*, Vol. 75, 1985, pp. 332-337.

Dufour, A. and R. F. Engle, Time and the Price Impact of a Trade, *Journal of Finance*, Vol. 55, 2000, pp. 2467-2498.

Engle, R.F. and J.R. Russell, Autoregressive conditional duration: a new model for irregularly spaced transaction data, *Econometrica*, Vol. 66, 1998, pp. 1127–1162.

Farrell, J. and C. Shapiro, Standard-Setting in High-Definition Television. *Brookings Papers on Economic Activity: Microeconomics*, 1992, pp. 1-93.

Garbade, K.D. and W.L. Silber, Technology, Communication and the Performance of Financial Markets: 1840-1975. *Journal of Finance*, Vol. 33, 1978, pp. 819–831.

Glosten, L. R. and P. M. Milgrom, Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneouly-Informed Traders. *Journal of Financial Economics*, Vol. 14, 1985, pp. 71-100.

Harris M. and A. Raviv, Allocations Mechanisms and the Design of Auctions. *Econometrica,* Vol. 49, 1981, pp. 1477-1499.

Hasbrouck, J. and G. Saar, Limit Orders and Volatility in Hybrid Market: The Island ECN. Working paper, New York University, 2004.

Iansiti M. and R. Levien, The Keystone Advantage: What the New Dynamics of Business Ecosystems Mean for Strategy, Innovation, and Sustainability. *Harvard Business School Publishing*, 2004.

Kraus, A. and H. Stoll, Price impacts of block trading on the New York Stock Exchange, *Journal of Finance*, Vol. 27, 1972, pp. 569-588

Kyle, A. Continuous Auctions and Insider Trading. *Econometrica* Vol. 53, 1985, pp. 1315-1335.

Malvey, P. F., C. M. Archibald and S. T. Flynn, Uniform-Price Auctions: Evaluation of the Treasury Experience. *Office of Market Finance, U.S. Treasury*, 1995.

Malvey P. F. and C. M. Archibald, "Uniform-Price Auctions: Update of the Treasury Experience. *Office of Market Finance, U.S. Treasury*, 1998.

Mendelson, H, Market Behavior in a Clearing House, *Econometrica*, Vol. 50, 1982, pp. 1505-1524.

Mendelson, H., Random Competitive Exchange: Price Distributions and Gains From Trade, *Journal of Economic Theory*, Vol. 37, 1985, pp. 254-280.

Mendelson, H. Pricing Computer Services: Queueing Effects. *Communications of the Association for Computing Machinery*, Vol. 28, 1985, pp. 312-321.

Mendelson, H., “Consolidation, Fragmentation and Market Performance.” *Journal of Financial and Quantitative Analysis*, Vol. 22, 1987, pp. 189-207.

Milgrom, P.R. and T. J. Weber, A Theory of Auctions and Competitive Bidding,” *Econometrica*, Vol. 50, 1982, pp. 1089-1122.

Saltzer, J. H., D. P. Reed, and D. D. Clark, End-to-end arguments in system design. *ACM Transactions on Computer Systems*, Vol. 2, 1984, pp. 277-288.

Reed, D. P., J. H. Saltzer, and D. D. Clark, Active Networking and End-To-End Arguments. *IEEE Network,* Vol. 12, 1998, pp. 69-71.

Santoli, M., The Big Squeeze: Rapid change on Wall Street is shrinking trading costs. The big winners: individual investors. *Barron's*, January 24, 2005.

The Standish Group, *Chaos Report*, 1995.

Van Ness, B.F., R.A. Van Ness and R. S. Warr, NASDAQ Trading and Trading Costs: 1993-2002, Working Paper, July 2004.