98-6
(66A)

EVALUATION OF ALTERNATIVE EXECUTION PROCEDURES IN FUTURES MARKETS

By

Yakov Amihud and Haim Mendelson

September 1, 1998

EVALUATION OF ALTERNATIVE EXECUTION PROCEDURES IN FUTURES MARKETS

By
Yakov Amihud and Haim Mendelson

TABLE OF CONTENTS

A. INTRODUCTION AND EXECUTIVE SUMMARY

On January 26, 1998, the Commodity Futures Trading Commission ("Commission") published a Concept Release regarding the Regulation of Noncompetitive Transactions Executed on or subject to the Rules of a Contract Market[1] ("Release"). The comment period on the part of the Release related to Alternative Execution Procedures ("AEPs") has been extended to September 1, 1998.[2] The Release sought comment regarding the possible effects of and appropriate standards for allowing AEPs, whereby certain phases of the trading process take place outside of the centralized, open and competitive marketplace. These issues are complex, as they require a careful analysis of the effect of AEPs on the quality of the market.

On August 11, 1998, the Commission held a public meeting regarding the application of the Cantor Financial Futures Exchange ("CFFE")[3] for designation as a new contract market for the trading of U.S. Treasury bond, ten-year note, five-year note, and two-year note futures contracts. While the CFFE application itself does not address this issue, at that meeting, questions were raised as to whether the core of the CFFE's trading system is an AEP that sacrifices the principle of an open and competitive marketplace by enabling a "workup" process similar to that used to facilitate block trades in the cash market for U.S. Treasury Securities.[4] Based upon our analysis, the trading rules of the proposed CFFE implement a novel AEP structure. If the Commission approves the CFFE application, it would, therefore, be adopting a new standard for the execution of noncompetitive transactions for futures contracts already traded in a mature, liquid contract market.

---

[1] CFTC Release # 98-006 (63 Fed. Reg. 3708).

[2] See CFTC Release #4137-98 dated April 27, 1998 (63 Fed. Reg. 24164).

[3] The CFFE application was first published on February 3, 1998 (63 Fed. Reg. 5505) and was since modified multiple times.

[4] See Section D.4, *infra*.

Given that the Commission is already considering the broader policy implications of AEPs, this report also addresses the desirability of approving the CFFE application within our economic evaluative framework.

The report sets forth an evaluative framework that builds on established principles and empirical evidence from many different markets. We discuss the tension between the desire to accommodate market participants who wish to consummate trades under trade execution structures that are less-regulated, and the harmful effects of market fragmentation that result from splitting the order flow. We show that a proper balance must take into account the effect of AEPs on market liquidity, price discovery (or informational efficiency) and market integrity. Market fragmentation reduces liquidity and increases overall trading costs, hampers price discovery and reduces the incentive to provide information to the market. Further, fragmentation undermines market integrity when trade prices can be worse than the best available bids and offers or when a class of traders receives inferior treatment.

We suggest that because many market participants prefer to trade on less restrictive markets, the futures markets risk a "race to the bottom," whereby markets successively match less-restrictive trading standards to prevent the loss of order flow. This calls for the establishment of market-wide AEP standards by the Commission prior to the commencement of trading on any market which features AEP.

We evaluate the CFFE using the above criteria. First, we show that the trading rules of the CFFE violate the requirements of an open and competitive market, i.e., the CFFE is an AEP. We show that the CFFE is fashioned after cash market procedures that are designed for the trading of large blocks. Because most of the trading volume in active U.S. Treasury Securities futures on the Chicago Board of Trade ("CBOT") is for orders of more than 50 contracts, this threatens the ability of the CBOT to continue to provide markets serving the public interests in efficient price discovery and risk management. Similar consequences may apply to any other U.S. futures exchange whose contracts CFFE chooses to mimic.

In evaluating the CFFE, we find it is likely to fragment the market for futures on U.S. Treasury Securities. We expect this will hurt liquidity and increase bid-ask spreads and search and delay costs. Because the CFFE will allow large orders to "trade through" superior bids or offers, market integrity and price discovery will be harmed. Further, the CFFE's trading rules discriminate against small orders, harming liquidity and market integrity. In addition, the system is unlikely to function well in turbulent times given its sequential execution procedure and the possibility that large orders will tie it up. We expect, however, that if the CBOT loses significant order flow, it will not have sufficient liquidity to serve as a "market of last resort" in times of stress or CFFE malfunction.

We expect the consequences of approving the CFFE application at this time to be a relaxation of market-wide trading standards, the possible loss of substantial order flow by the CBOT, a less liquid market for futures on U.S. Treasury Securities, less efficient price discovery, an increased cost of hedging U.S. Treasury Securities and a higher cost of debt for the U.S. Government.

The structure of this report is as follows. Section B reviews basic criteria for evaluating a trading market. Section C studies the issue of market fragmentation. Sections D and E consider three AEP case studies: The NYSE large block trading procedure, the interdealer market in U.S. Treasury Securities, and the CFFE.

B. MARKET QUALITY: PRINCIPLES AND EVALUATION CRITERIA

1. Futures Markets

(a) Introduction

Futures contracts traded on exchanges are agreements to buy or sell a standardized[5] commodity at some time in the future. Futures exchanges play an important role in the functioning of the U.S. economy. First, they enable economic agents (such as farmers, investors and traders) to hedge against changes in the value of the underlying commodity. Second, they provide a superior vehicle for the discovery and dissemination of price information for the underlying commodity. Whereas the risk management role of futures contracts could be accomplished by customized contracts for forward delivery between the counterparties, such contracts are costly, illiquid, and do not provide efficient pricing. Futures contracts have the advantages of liquidity, competitive pricing, convenience and low cost, which account for their popularity and for their extensive use by agents in the economy.

A futures contract on an underlying asset serves to hedge against changes in the price of the underlying asset (or a related asset) in which a person is holding a position or about which he has a commitment. The purpose of holding a hedged position is to reduce exposure to the risk inherent in the position or commitment. For example, an investor who holds a U.S. Treasury Bond can hedge against interest-rate risk by taking a short position in T-Bond futures contracts. If interest rates rise, the value of the bond will decline and the investor will lose on his cash market position, but the short position in the futures contract will rise in value, generating a gain that can offset the cash-market loss to any desired extent. Thus, a properly designed hedged position can lock in the value of the T-Bond and reduce the investor's risk. Or, suppose an investor plans on buying a T-Bond in, say, a month, when he expects to receive some cash inflow. The investor does not have the money now, but wants to lock in the value of the bond that he will buy; put differently, he wants to ensure that he receives the current interest rate. This investor can buy (long) a T-Bond futures contract that will rise in value if interest rates fall, and compensate him for the higher price he will have to pay for the T-Bond when he receives the money a month later. Farmers and others in the agricultural marketing chain can similarly reduce the risk of price change in agricultural commodities by hedging in the futures markets.

Clearly, sophisticated investors carry far more complex hedged positions. Financial institutions, like banks, that borrow short-term funds and make long-term commitments, need to hedge against changes in the slope of the yield curve and against the difference in duration of their assets and liabilities. The futures markets enable them to do that in a convenient and low-cost way, and to unwind their positions at any time by simply selling or buying futures contracts in the market.

The ability to hedge at low cost, which is made feasible by the futures markets, reduces the risk of owning or using the underlying commodity and encourages risk-averse agents to hold a position or take a commitment in it. The ease and convenience of managing risk with futures contracts facilitate shifting the risk to speculators, who are better able or more willing to bear it, and enable economic agents to invest in the commodities of their choice without undue exposure to risk. One may think of the ability to hedge as the ability to buy price-insurance in a flexible and low-cost way.

---

[5] These agreements are standardized according to the quality, quantity, delivery time and location of the underlying commodity.

(b) Value of Futures Markets

Futures markets are important since they make futures contracts liquid. Well-functioning futures markets can also improve the quality of the market where the underlying commodity is traded. Bessembinder and Seguin[6] examined the effect of the introduction of stock index futures on stock market volatility and volume. They found that the contemporaneous spot market volume/volatility coefficient declined significantly in the period since equity futures trading commenced, as compared to earlier years. They further found that equity volatility declines as predictable futures trading activity increases. Their results support the proposition that futures trading increases the depth of the market for the underlying, enhances its liquidity and reduces its volatility. These results are consistent with the theoretical findings of Grossman and Miller[7] and Grossman[8], who showed that the existence of stock index futures and the arbitrage between the futures and the underlying stocks increases the liquidity and decreases the volatility of the cash market. These results are also consistent with the earlier empirical findings by Edwards[9], who showed that the introduction of stock index futures reduces equity volatility.

Liquid futures markets create the related advantage of efficient price discovery and dissemination. And, when a futures contract is more liquid than the underlying commodity, it serves to quickly discover and disseminate information about that commodity, facilitating and expediting the price discovery process for the underlying.[10] For example, the futures contract on

---

[6] H. Bessembinder and P. J. Seguin, "Futures Trading Activity and Stock Price Volatility," *Journal of Finance* 47, 1992, 2015-2034.

[7] S.J Grossman and M. Miller, "Liquidity and Market Structure," *Journal of Finance* 43, 1988, 617-633.

[8] S.J. Grossman, "An Analysis of the Implications for Stock and Futures Price Volatility of Program Trading and Dynamic Hedging Strategies," *Journal of Business* 61, 1988, 275-298.

[9] F.R. Edwards, "Futures Trading and Cash Market Volatility: Stock Index and Interest Rate Futures," *Journal of Futures Markets* 8, 1988, 421-439.

[10] See K. D. Garbade and W. L. Silber, "Price Movements and Price Discovery in Futures and Cash Markets," *Review of Economics and Statistics* 65, 1983, 289-297.

the S&P 500 stock index is more liquid than the underlying: taking a position in the index requires trading 500 different stocks, whereas a single transaction is sufficient to buy the futures contract on the index. Thus, when an investor obtains information about the stock market, it is much faster to act on this information by trading the futures contract than by trading the underlying 500 stocks. Also, while the average bid-ask spreads and brokerage fees for the stocks in the S&P 500 index range between 0.5% to 1.0%, the transaction costs of the respective futures contract are a fraction of this amount. New information can therefore be incorporated into the futures price more quickly than into the price of the underlying index. This implies that the speed of adjustment of prices to new information is faster in the futures market.[11]

Empirical evidence shows that information is indeed more quickly incorporated in the futures prices than in the prices of the underlying commodities. Kawaller, Koch and Koch[12] examined whether the S&P 500 futures contract serves as an effective price discovery vehicle for the S&P 500 stock index. They found that futures prices lead stock prices, and that this lead extends between 20 and 45 minutes, while the lead from stock prices to futures prices rarely extends beyond one minute. Stoll and Whaley[13] also found that stock index futures lead the stock index. A recent study of stock index futures by Pizzi, Economopoulos and O'Neill[14] found that futures markets lead the spot market by at least 20 minutes, whereas the lead of the spot market

[11] See analysis and discussion in Y. Amihud and H. Mendelson, "Index and Index-Futures Returns," *Journal of Auditing, Accounting and Finance*, 1989, 415-431.
[12] I. Kawaller, P. Koch and T. Koch, "The Temporal Price Relationship Between S&P 500 Futures and the S&P 500 Index," *Journal of Finance* 42, 1987, 1309-1329.
[13] H. R. Stoll and R. E. Whaley, "Volatility and Futures: Message versus Messenger," *Journal of Portfolio Management*, 1988, 20-22.
[14] M. A. Pizzi, A. J. Economopoulos and H. M. O'Neill, "An Examination of the Relationship between Stock Index Cash and Futures Markets: A Cointegration Approach," *Journal of Futures Markets* 18, 1998, 297-305.

over the active futures markets is three minutes. Ederington and Lee[15] found that the price adjustment to information release in the futures market occurs within one minute of the release. Some futures markets are more efficient in the dissemination of information than other markets. For example, Ederington and Lee found that information is disseminated faster in the T-Bond futures market than in the Eurodollar and Deutsche Mark futures markets.

In summary, futures markets both reduce the cost of taking a position and improve efficiency in the market for the underlying. As a result, their liquidity is important for the economy as a whole.

(c) Futures Markets vs. Securities Markets and Product Innovation

Most of the results in market microstructure theory apply to any trading market, regardless of whether the financial instruments traded in the market are stocks, debt instruments or futures contracts. Indeed, the basic relationships governing the economics of markets are not highly sensitive to the nature of the traded instruments, which in most cases affects the *parameters* of the models rather than their fundamental structure. Empirical tests of market microstructure theories indeed corroborate that the same fundamental principles govern the different types of trading markets.[16]

This is not to say that the same policy considerations should also govern securities and futures markets. Differences such as the fact that futures contracts do not require an initial public offering and their supply is unlimited leads to significant regulatory differences. Similarly,

[15] L. H. Ederington and J. H. Lee, "How Markets Process Information: News Releases and Volatility," *Journal of Finance* 48, 1993, 1161-1190.

differences in the volatility of the instruments and in the pace of trading have important implications on the value of timeliness, the cost of delay and the optimal market structures that ensue. Further, the different statutory schemes giving rise to the regulation of securities and commodity futures mean that the history and evolution of futures markets followed different paths from their securities markets counterparts.

In particular, whereas the U.S. futures markets have continually evolved as centralized auction markets, the securities markets have always been a mix of auction markets and over-the-counter markets. Correspondingly, the futures markets have consistently adhered to a strict interpretation of the concept of an open and competitive market, historically requiring all futures transactions to be negotiated and executed in an auction market, except for EFPs. In contrast, U.S. securities regulation historically allowed, and sometimes encouraged, over-the-counter trading of securities and unlisted trading privileges.[17] As a result, U.S. securities regulators had to adopt countermeasures that were designed to prevent or at least mitigate the harmful effects of market fragmentation discussed in Section C, *infra*. Thus, while AEPs are only a recent phenomenon in the futures industry, their securities markets equivalents have existed over extended periods of time.

As a result, there is a rich body of empirical evidence on the effects of different AEP alternatives based on the experience of U.S. and some foreign securities markets. In interpreting this evidence, it is important to extract the valuable lessons that apply to any trading market while keeping in mind that the ultimate regulatory action need not be the same due to the different nature of futures markets. Some of these general lessons relate to the effects of market fragmentation on liquidity, price behavior, the impact of different block trading procedures, the incentives to provide

[16] This is reflected in the work cited later in this report.

liquidity and the effects of AEPs on market efficiency. Indeed, these lessons are corroborated by the more limited evidence from the futures markets (primarily from Europe). However, given these general lessons, public policy should fully take into account the differences between the futures and securities markets that may lead to sharply different actions. We believe a key difference between securities and futures markets is in the nature of innovation.

For most securities markets, the most important innovations are in the domain of trading technology. The securities themselves are issued by an independent issuer and, once listed, they are often traded in multiple markets. The market that lists the security certainly does not have to "develop" it — the security needs to be listed and traded following the common trading procedures of the securities market. Once the security is listed, other securities exchanges may trade the security following a straightforward procedure that provides them, almost automatically, Unlisted Trading Privileges in the security.[18]

In contrast, futures markets innovate on two different dimensions: product development and trading technology.[19] Futures contracts are innovative and complex products that must be developed by the exchange itself. Rather than being driven by tax reasons, as is the case for a number of other innovations in the financial services industry, futures contracts belong to the class of financial innovations that are driven by transaction cost savings and improvement in liquidity.[20] As a result, the innovation inherent in financial futures leads to real economic benefits. Like any other substantive innovation, new futures contracts require major investments in research and development and in implementation, and are subject to a substantial risk of failure.

---

[17] This approach was not followed by all securities regulators worldwide, and non-U.S. regulatory schemes often favor centralized auction markets for securities as well as futures.
[18] See 5 U.S.C. §78l(f).
[19] On the nature of financial innovation in futures markets, see M. H. Miller, *Financial Innovations and Market Volatility*, Cambridge, MA: Basil Blackwell, 1991.
[20] See R. C. Merton, "The Financial System and Economic Performance," *Journal of Financial Services Research* 4, 1990, 263-300.

Black[21] studied a sample of 26 new financial futures contracts, finding that only nine of them were successful. Her criterion for success was having the contract listed in *The Wall Street Journal* within three years of the contract start date.[22] Using a similar criterion, Carlton[23] reports that a large fraction of contracts cease being reported in *The Wall Street Journal* within two years of their initial listing. Sandor[24] reports that of the 56 new contracts introduced by the various futures exchanges between 1960 and 1970, eighteen were successful. The failure rate is similar in the nineties: only a third of futures contracts approved by the Commission over the 1991-95 period remained listed on an exchange by April of 1996.[25]

Like any innovative product, the success of futures contracts depends on skillful design, experimentation, nurturing and investment of resources. In his analysis of innovation in the futures markets, Silber[26] pointed out that many aspects of contract design that seem trivial at the outset may ultimately prove crucial and determine the success of the contract. An example is the rise and fall of the futures contract on Government National Mortgage Association Collateralized Depository Receipts (GNMA CDR), introduced in 1975 by the CBOT.[27] It traded 20,125 contracts in less than three months in its year of introduction, its trading volume peaked at 2,326,292 contracts in 1980, and in 1987 it traded only 7,583 contracts (see also Duffie[28]). The reason for the failure was a flaw in the design of this futures contract: the delivery options included in the contract

---

[21] D.G. Black, *Success and Failure of Futures Contracts: Theory and Empirical Evidence*, Salomon Brothers Center Monograph Series in Finance and Economics, 1986, Monograph 1986-1.

[22] Listing in *The Wall Street Journal* required both a daily trading volume of 1,000 contracts and open interest of at least 5,000 contracts.

[23] D.W. Carlton, "Futures Markets: Their Purpose, Their History, Their Growth, Their Successes and Failures," *Journal of Futures Markets* 4, 1984, 237-271.

[24] R.L. Sandor, "Innovation by an Exchange: A Case Study of the Development of the Plywood Futures Contract," *Journal of Law and Economics* 16, 1973, 119-136.

[25] S. McGee, "Exchanges Race to Offer New Products," *The Wall Street Journal*, April 8, 1996, C1.

[26] W. Silber, "Innovation, Competition, and New Contract Design in Futures Markets," *Journal of Futures Markets* 1, 1981, 123-155.

[27] See analysis in E. Johnston and J. McConnell, "Requiem for a Market: An Analysis of the Rise and Fall of a Financial Futures Contract," *Review of Financial Studies* 1988, 1-23.

[28] D. Duffie, *Futures Markets*, Prentice Hall, Englewood Cliffs, NJ, 1989, at 339-342.

reduced its hedging effectiveness. Duffie proposed: "It appears that the formula determining the discount on delivery substitutions caused the futures price to be relatively poorly correlated with the spot price of newly issued mortgages. Since newly issued mortgages are felt to be the main source of hedging demand, the contract design in effect caused the GNMA futures contract to be a relatively poor hedge" (p. 340). The decline in the hedging interest in the GNMA CDR contract led to a decline in its trading volume and to its demise.

In addition to the uncertainty regarding the success of a new futures contract, the cost of development is substantial. *The Wall Street Journal*[29] reports that the cost of developing a new futures contract is at least a few hundred thousand dollars, and in the case of the New York Mercantile Exchange's electricity futures contract, the cost was about $2 million. Silber[30] estimated that the *direct* cost of developing a new futures contract during the period 1970-1979 was $70,000, which, in today's terms, would be in excess of $200,000.[31] However, Silber pointed out that "the direct cost components are only the tip of the iceberg" (p. 131). The actual cost is much higher, involving expenses such as the cost of time and attention of exchange staff and members, promotion and education. Silber also estimated that it could take as much as a year or two to develop a new product such as the currency futures introduced in 1972 by the Chicago Mercantile Exchange and the financial futures introduced in 1975 by the CBOT.

Sandor[32] described the development process of the CBOT's plywood futures contract. The process is similar to that of other risky innovations, starting with a formal examination to determine whether or not the commodity can be adapted to futures trading; moving into preliminary contract design; proceeding with detailed design in consultation with exchange staff, members and other

---

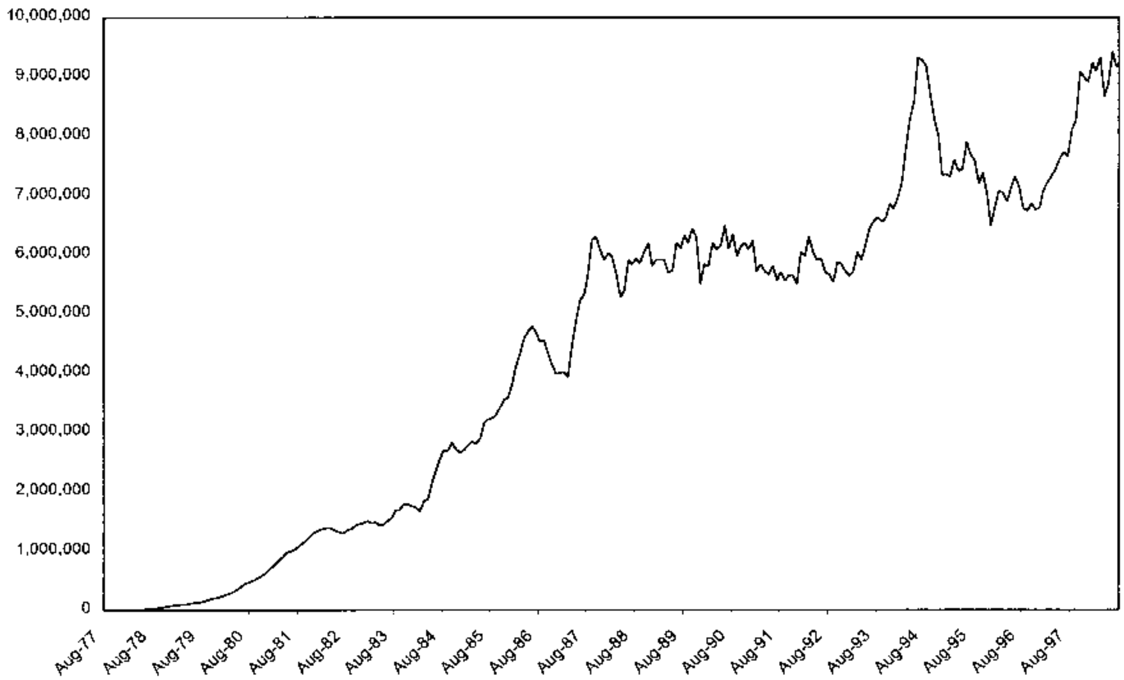
[29] McGee, *Op. Cit.*
[30] Silber, *Op. Cit.*
[31] Taking into account the rise in the hourly wage rate between 1975 and 1998.
[32] Sandor, *Op. Cit.*

market participants and finally moving into product introduction and the commencement of trading. The development process continues beyond the product introduction, involving changes in product specifications to broaden contract appeal.

Indeed, the introduction of a new futures contract requires an investment in educating potential users, attracting interest in the product and, most importantly, nurturing the product in the marketplace to develop sufficient liquidity so it will survive and thrive. The development of liquidity is a long process that can take literally years. Figure B-1 shows the monthly trading volumes[33] in the CBOT's most successful product, the Treasury Bond futures, since their introduction on August 22, 1977. The contract volume grew gradually over a ten-year period, as liquidity increased, use of the product became more prevalent and the hedging demand satisfied by the contract increased over time.

[33] To smooth out some of the volume fluctuations, the figure shows 6-month moving averages (for each month, the number shown is the average monthly volume for the half-year period ending with that month).

Figure B-1: Monthly Contract Volume - CBOT Bond Futures



The considerable resources required to develop a successful new futures contract and the significant risk of failure create strong incentives for futures contract imitation. That is, a futures exchange may want to clone a contract that was introduced by another exchange, passed the market test and was proven successful. In addition to avoiding the risk of design failure by cloning or imitating only winners, the direct cost of developing a competitive modification was estimated to be about one-third of the cost of developing a new contract.[34] The indirect costs associated with promotion and educating the public are far lower, given that the imitator has picked a proven winner that has already been accepted by the public. This represents free-riding by the imitating

[34] Silber, *Op. Cit.*

exchange on the resources invested and the risk borne by the exchange that had initiated and developed the successful futures contract. Having unfettered imitation of successful futures contracts reduces the incentives of exchanges to invest in the development of new futures contracts and hampers innovation.

Free-riding can result in *under*-investment in the development of new futures contracts, because the developer loses part of the trading revenues it had hoped to recoup to a competing exchange. The problem is similar to that of protecting other intellectual property rights: the development costs (including the expected cost of failure) are fixed and sunk by the time the futures contract is successfully traded. Naturally, the innovating exchange will have to price its services above short-run marginal cost to recover its development costs. Pricing at short-run marginal cost will lead to the demise of investments in innovation and new product development. Yet, pricing above short-run marginal cost and the appearance of current profit (needed to recover past development costs and to finance new product development efforts) attracts other exchanges to imitate the winning contract while pricing their services lower. While in the short run, regulators and the public might welcome such cost reduction, it can hurt innovation in the long run: if the innovator bears the costs of product development without receiving commensurate benefits, the incentive to innovate is eliminated.

This issue is related to the structure of futures markets as follows. Similar to other innovations, the innovating exchange has a first-mover advantage because it was the first to develop liquidity in the new contract. This liquidity advantage is significant, because order flow tends to attract additional order flow,[35] and it serves to compensate the innovating exchange for its research and development effort. Accordingly, the structure of the futures markets has evolved so that while cloning and imitation are common practice, futures exchanges are capable of partially protecting the liquidity they developed in their innovative products, thus being rewarded for their innovation. The result is sustained innovation in the futures industry. For example, McGee[36] reports that in the first three months of 1996, the Commission approved 42 new futures contracts, compared to eight in the first three months of 1995 and 59 in all of 1995. This rate of innovation, however, depends on the exchanges' ability to reap the rewards when they are successful.

In the securities markets, the Securities and Exchange Commission ("SEC") consistently encouraged a combination of multiple trading[37] of securities and electronic market linkages.[38] The former are intended to foster competition, and the latter to prevent the associated market fragmentation. For example, the Intermarket Trading System makes the best quotes in each market center available to the other market centers.[39] This results in a practice of "quote matching," where a "satellite" exchange can make the quotes of the best market center available in its own market by simply matching them. For New York Stock Exchange ("NYSE") listed stock, for example, the best quotes were provided in most case by the NYSE and were matched by competing regional exchanges, making much of the liquidity on NYSE available to its regional competitors.[40] We have argued before that this form of free-riding hurts liquidity in the securities markets.[41]

In the futures markets, the effect of similar linkages would be much more damaging. Not only would they hurt liquidity, they would also deprive innovating exchanges of the rewards to their innovation. As discussed above, the liquidity advantage of an innovating exchange currently provides it an incentive to innovate. If competing market centers could free-ride on that liquidity,

---

[35] Formally, this is a reflection of the positive *liquidity externality* discussed in Section B.3(a), *infra*.
[36] McGee, *Op. Cit.*
[37] That is, trading the same security in multiple markets.
[38] See, Y. Amihud and H. Mendelson, "A New Approach to the Regulation of Trading Across Securities Markets," *New York University Law Review* 71, 1996, 1411-1466; Y. Amihud and H. Mendelson, "How (Not) to Integrate the European Capital Markets," *European Financial Integration* (A. Giovannini and C. Mayer, editors), Cambridge University Press, 1991, 73-100.
[39] See, Y. Amihud and H. Mendelson, "How (Not) to Integrate the European Capital Markets," *Op. Cit.*
[40] See Section C, *infra*.

the innovator's liquidity advantage would disappear, and so would the incentive to develop new products. Thus, in evaluating market structure solutions for the futures markets, it is important to consider the effects on the incentives to engage in product innovation — a consideration that does not come into play in most securities markets.

In summary, the problems caused by having a futures contract trading in multiple markets are more complex than those associated with a standard security – such as a stock – traded in this way. As with any security traded in multiple, separate markets, there are problems of market fragmentation that reduce market quality; see Section C, *infra*. The experience of securities markets is highly informative, and shows the problems that arise. In futures markets, there is another important consideration that does not arise in most securities markets: imitation by and trading on a different market reduces the incentives to invest in new product development. Hence, market structure solutions that facilitate duplicating not only a contract but also its liquidity are undesirable.

2. Evaluating Trading Markets

A trading market (such as an exchange) is a mechanism that serves to communicate orders for trading in standardized contracts or securities and to transform these orders into trades, specified by the quantities exchanged between the counterparties and by the transaction prices. The trading process consists of three phases: *(i)* information assimilation, which leads to the creation of orders; *(ii)* order communications to the exchange; and *(iii)* order execution, which converts the orders into

[41] Y. Amihud and H. Mendelson, "How (Not) to Integrate the European Capital Markets," *Op. Cit.*

trades. Depending on market design, the last two phases can generate new information that feeds into the information assimilation phase and creates new orders and trades.

At the heart of the trading market is a communication system, an execution system and the market's operating rules. The communication system helps communicate and display orders, quotes, prices and other messages. The execution system transforms orders into trades. The market's operating rules govern the operations of the two systems and specify which communications are permissible and feasible, how orders are expressed, and how they are transformed into trades. In an electronic exchange, the market operating rules require both the communication system and the execution system to be electronic. Thus, once an order is communicated to the exchange, it can be executed with no further human intervention.

Electronic exchanges have become viable due to developments in information technology, globalization, and the experience gleaned from the successes and failures of electronic contenders in the past. An increasing number of futures markets have adopted electronic trading procedures, and their number and importance will continue to increase.[42] However, mere use of electronics does not guarantee market effectiveness; automation can reduce certain trading costs, but market quality is at least as important. For example, the National Securities Trading System launched as part of the Cincinnati Stock Exchange in 1978 was touted as the electronic stock exchange of the future, but it never attracted meaningful order flow. A market's operating rules are therefore at least as important as the technology used to implement them. What ultimately matters is the *overall* performance of the trading market.

[42] For a discussion of automation in futures markets, see Section B.5, *infra*.

How is that performance measured? In our earlier work,[43] we set out three key characteristics[44] for evaluating the quality of a trading market:

(a) Liquidity,

(b) Informational Efficiency, and

(c) Market Integrity.

In addition, *product innovation* is an important criterion for evaluating futures markets. Whereas the above criteria for the evaluation of trading markets take the instrument being traded as given, futures markets develop both the instrument and the mechanism for trading it. This criterion and its interaction with the market structure was already discussed in Section B.1(c). We discuss each of other three criteria below.

(a) Liquidity

The key economic function of a trading market is to provide liquidity,[45] i.e., to enable traders to buy and sell the traded instruments *quickly* and at *low cost*. Trading markets increase economic efficiency by reducing the cost and increasing the speed of search for compatible trading partners. The quality of the market depends on investors' *overall* costs, which take into account the execution price, the direct costs of execution and the opportunity cost of delaying a trade. These direct and indirect illiquidity costs are described below.

---

[43] See, Y. Amihud and H. Mendelson, "Trading Mechanisms and Value Discovery: Cross-National Evidence and Policy Implications," *Carnegie-Rochester Conference Series on Public Policy* 34, 1992, 105-134; Y. Amihud and H. Mendelson, "How (Not) to Integrate the European Capital Markets," *Op. Cit.*; Y. Amihud, H. Mendelson and B. Lauterbach, "Market Microstructure and Securities Values: Evidence from the Tel Aviv Stock Exchange," *Journal of Financial Economics*, 45, 1997, 365-390.

[44] As discussed below, the three major characteristics aggregate a number of sub-characteristics.

[45] See Y. Amihud and H. Mendelson, "Liquidity, Asset Prices and Financial Policy," *Financial Analysts Journal* 47, November/December 1991, 56-66.

(1) Bid-Ask Spread and Market-Impact Costs

The high equilibrium price volatility in the underlying commodities is a key reason for the development of futures markets. Because of the high frequency of price changes, virtually all futures markets are continuous markets.[46] A continuous market enables market participants to trade a small order instantaneously at the market bid (for a sell order) or ask (for a buy order) price. In over-the-counter dealer markets, the bid and ask prices are posted on computer screens or quoted over the phone. In auction markets, bid and ask prices are either quoted by market participants on the floor of an exchange (in the case of a physical floor) or on a limit order book (in the case of an electronic exchange), where the bid and ask prices represent the best limit price to buy and sell, respectively.

The difference between the bid price and the ask price—the bid-ask spread—represents a cost of illiquidity. A high-quality market should provide for a narrow bid-ask spread, which implies a lower cost for an instantaneous "round-trip" transaction.

However, the bid and ask quotes are good for limited quantities only. If a trader wants to sell a larger quantity, he will likely end up obtaining for the whole quantity a lower average price than the best quoted bid price.[47] That is, the sale of a large quantity likely has a negative impact on the sale price, and on average, the negative impact is an increasing function of the amount sold.

Thus, the sale of a large block typically causes a price decline, of which part is temporary. For example, the temporary market impact costs of block sales were estimated at 1.86% for NYSE and American Stock Exchange ("AMEX") transactions, and 3.28% for National Market System

[46] Securities markets are often organized as periodic clearing markets (see, *e.g.*, a review of the various markets that operate in this way in R. A. Schwartz, *Reshaping the Equity Markets*, Business One Irwin, 1993). However, because futures markets are almost always continuous markets, we focus on this case. See, however, Section E.6.
[47] Analogously, a large buy order will typically make the buyer pay for the entire quantity a higher average price than the best quoted ask price.

("NMS") stocks on the NASDAQ Stock Market.[48] The overall decline in the price received by the seller of a large block, or the price paid by the buyer of a large block, is called market-impact cost. Block traders are naturally concerned about the market impact of their trades, which can result in significant costs.

A high-quality market has *depth*, i.e., the quantity that can be traded at the bid and ask prices is large. The deeper the market, the lower the market-impact costs. A deep market has greater *resiliency*: the arrival of an unexpectedly large buy or sell order will cause a minimal price impact. In a *thin* market which is lacking in depth, i.e., in a market where the quantity that can be sold or bought at the quoted bid or ask prices (respectively) is small, the market impact cost is greater. Thus, market depth has an important role in reducing the cost of immediacy and providing liquidity.

(2) Search and Delay Costs

The buyer or seller of a large block faces a tradeoff between market impact costs and execution delays. When this is allowed, a trader may search "upstairs" for better prices and approach others (typically through a block positioning intermediary) to solicit orders on the other side of the market. However, the delay associated with this search process can be costly, because an unfavorable price change can take place between the time the trader made his initial trading decision and the time of ultimate execution. For example, suppose the trader decides to short a contract whose price is expected to decline. By delaying the execution, the trader is exposed to the risk that the price will decline while the search for a better price is still under way. Worse yet, the search process itself may alert the rest of the market to the impending sale of a large block, moving the market down—

[48] These numbers pertain to large blocks executed in the "upstairs" market; see Section D, *infra*. See D. B. Keim and A. Madhavan, "The Upstairs Market for Large-Block Transactions: Analysis and Measurement of Price Effects," *Review*

the very outcome the trader wanted to avoid. Then, the delay in execution makes the trader incur a loss.

While delay costs are difficult to measure, they can be substantial, especially in fast-moving markets. In a study of execution costs for equity trades by pension funds in the second and third quarters of 1997, Wagner[49] found that delay costs accounted for about half of the total cost of order execution.[50] Delay costs are high in futures markets, where speed in execution is paramount. Grossman and Miller[51] observe that futures markets arise to accommodate the sustained demand for quick hedging. Hence, "Successful futures markets are the leading examples of markets where the demand for immediacy is high,"[52] and futures markets are successful exactly for those commodities where price volatility, and hence the risk of delaying trading, is high.

A high-quality market should minimize these delay costs by enabling speedy executions. This can be achieved in an auction market with a multitude of participants, e.g., locals who stand ready to absorb incoming orders. Then, the search cost of contacting potential traders on the other side (e.g., buyers when the order is to sell) is low and the delay is minimized.

---

of Financial Studies 9, 1996, 1-36.

[49] Wayne Wagner, Plexus Group, March 1998.

[50] Specifically, the delay cost was 72 basis points compared to a total of 147 basis points (72/147) for Small-Cap Growth funds, 63/123 for Small Cap Value funds, 61/96 for Index funds, 32/63 for Large-Cap Growth funds and 13/36 for Large-Cap Value funds.

[51] S.J. Grossman and M. H. Miller, *Op. Cit.* at 617-633.

[52] *Ibid.*

(3) Execution Risk

A trader may be exposed to *execution risk* if the trading system does not guarantee an execution within the desired period of time. A market order—an order to buy or sell at the best obtainable price—is designed to ensure immediate execution. However, this requires that the market's operating rules ensure immediate execution, and that the market has sufficient liquidity so there is a counterparty willing to execute the order on the other side. The execution risk of a market order should be nil if the market is liquid and the trading system is designed to execute it promptly. If a trading system has a built-in delay between the time an order is entered and the time it is executed, it exposes the trader to an execution risk. Then, even a market order exposes the trader to the risk that the market will move against him, e.g., the price will rise during the time that elapses between the time a buy order is entered and the time it is executed.

(4) Commissions and Fees

These reflect additional direct costs of trading, including brokerage commissions, exchange fees, taxes and other direct costs. Because traders are interested in their *total* cost of buying or *net* proceeds from selling, high commissions and exchange fees do not necessarily imply that the trader is worse off paying them, since there may be a tradeoff between these direct costs and other illiquidity costs. For example, a block trader may prefer a market with higher commissions and fees if trading there results in lower market-impact costs. Chan and Lakonishok[53] found that for institutional investors' sales of large equity blocks, market impact costs were negatively related to the commission expense: the higher the commission, the lower the market impact cost.

[53] L.K.C. Chan and J. Lakonishok, "The Behavior of Stock Prices Around Institutional Trades," *Journal of Finance* 50, 1995, 1147-1174.

Institutional investors were therefore willing to pay higher commissions in return for lower market-impact costs.

(b) Informational Efficiency and Price Discovery

An important role of a trading market is the concentration of information. A centralized market provides a mechanism that aggregates information from a variety of sources into the prices of traded contracts. Within the multitude of traders who participate in the market, each comes with a single piece of information, partly being common, and partly reflecting privately-produced information. A central competitive auction market aggregates the individual pieces of information to produce a consensus in the form of a market price. The market price summarizes the individual bits and pieces of information and provides an important service to anyone who uses the observed price in his decision-making process.

The more central an instrument in the economy, the more important its informational role. For example, the information on futures contracts on U.S. Treasury Securities is extremely valuable because of the central role of interest rates in the economy. In addition to their role in macroeconomic decisions, the information embedded in the prices of these contracts is important for traders in all financial and commodity markets. Expectations about interest rates are an important input into almost any economic decision, and those expectations depend on price signals that reflect all available information.

An efficient market reflects all available information, improving price discovery and reducing price uncertainty. Price uncertainty unnecessarily increases the risk of trading and the implicit costs of hedging. A quality market mechanism should minimize the uncertainty that results from noise, or transitory price changes around equilibrium values.[54] Changes in information that affect the value of futures contracts should be fully and promptly incorporated into the futures price. Price uncertainty that results from an inefficient trading mechanism confuses market participants, makes it harder for decision-makers to discern information from random fluctuations in the prices they observe, and slows down the price discovery process.

Further, price uncertainty increases the costs of illiquidity. In the futures markets, locals who take a position and provide liquidity to the market are not fully diversified. Being risk-averse, they demand a premium for bearing the risk inherent in the positions they take. The greater the price uncertainty and the less efficient the market, the greater that premium. Hence, price uncertainty increases the costs of illiquidity.

Evidence for stocks shows[55] that the greater the uncertainty associated with price changes for a security, the wider its bid-ask spread—that is, *price uncertainty reduces liquidity*. The effect of price uncertainty on transaction costs is strongly present in the futures markets, as shown by Wang, Yau and Baptiste.[56] They found that, after controlling for other factors, the bid-ask spread on both commodity futures and financial futures was highly sensitive to the intraday price volatility. The elasticity of the bid-ask spread with respect to the intraday price volatility was 0.478 for copper futures, 0.377 for gold futures, 0.326 for T-Bond futures, 0.306 for Deutche Mark futures, 0.262 for wheat futures and 0.204 for soybean futures.[57] This means, for example, that for

[54] This is to be distinguished from the volatility of the *value* of the underlying asset, which should be fully reflected in the futures price.
[55] See G. J. Benston, and R. L. Hagerman, "Determinants of the Bid-Ask Spreads in the Over-the-Counter Market," *Journal of Financial Economics* 1, 1974, 353-364. Benston and Hagerman measured price uncertainty by examining the *idiosyncratic* (or unsystematic) variance associated with price changes in the security.
[56] G.H.K. Wang, J. Yau and T. Baptiste, "Trading Volume and Transaction Costs in the Futures Markets," *Journal of Futures Markets* 17, 1997, 757-780.
[57] When the elasticity is equal to *e*, this means that if price volatility increases by one percent relative to its base value, the bid-ask spread will increase by *e* percent relative to its base value.

the T-Bond futures, a 1% increase in price volatility (relative to its base value) will increase the bid-ask spread by about 1/3%. Thus, increased price uncertainty reduces liquidity.

In a central auction market, the participation of many traders in the transaction process ensures that random, idiosyncratic price fluctuations will be quickly corrected and eliminated. Thus, the resulting prices provide less noisy signals about the fundamental information regarding interest rates and commodity values. As discussed in Section C, *infra*, if trading is scattered among a number of trading venues, price uncertainty increases. A quality market system should facilitate the process by which market participants can find the best price at which to execute their orders. This can generally be provided by having a *single* place where this information is available. In addition, a high-quality market should provide traders with an assurance of execution at a single best price at any given point in time.

Informational efficiency is closely related to liquidity. First, larger bid-ask spreads imply more price fluctuations between the bid and ask prices, leading to greater price uncertainty. Second, the price impact of a large block transaction increases price uncertainty because it is largely transitory, and prices are at least partially reversed after the initial impact. Further, *trade-throughs,* which occur when orders are executed at prices that are inferior to the best bid and offer available in the market, lower both liquidity and efficiency. From the perspective of liquidity, a trade-through means that the trader did not receive "best execution," thereby increasing her cost of execution. From the perspective of market efficiency, a trade-through means that transaction prices do not accurately reflect the state of the market, i.e., market prices are unreliable indicators of the information available in the market and the market is informationally inefficient.

(c) Market Integrity

Market integrity is another key feature of a trading market. Doubts regarding the integrity of a market can lead at the very least to inefficiencies, and at their worst to the ultimate collapse of the market. Even the slightest doubt regarding the honesty of a trading market can lead to extremely harmful consequences. Market participants may then withdraw from the market or trade with great caution, rendering the market illiquid and undermining its function as the provider of price information.

Market integrity has a number of dimensions. First, functional integrity means that the market functions properly, namely the market communication system has the necessary capacity and the trading system can guarantee sufficient throughput. Functional integrity is measured during periods of stress and requires that the market will not collapse and that it has adequate capacity and throughput necessary to handle peak order flows and trading volumes. Functional integrity depends both on the technical infrastructure and on the market's trading rules. Trading rules are inconsistent with functional integrity where they are designed in a way such that the market cannot withstand stress or makes market operation depend on some other "market of last resort."

Another dimension of market integrity is equal treatment of all market participants. When a class of market participants is at a disadvantage compared to others, market integrity suffers. In reaction, the disadvantaged market participants may become less active or altogether withdraw from the market, leading to lower liquidity.

For example, if one class of market participants has a lower priority of execution compared to another class, they cannot trade as quickly on new information as their market competitors. Given the zero-sum-game nature of futures markets, this means that on average, they will lose money to the class that has priority of execution. As a result, they will reduce their propensity to trade or will stay out of the market altogether.[58] This reduces the liquidity of the futures market overall, and can increase the effective cost of participation in related markets.

Another important dimension of market integrity relates to the market's informational efficiency. Market participants expect the market to follow the "law of one price" which requires that at a given point in time, there is a single market price. Further, the market price should be determined in an honest, open and competitive process. This helps traders and regulators enforce "best execution" — another hallmark of market integrity. In addition, open and competitive pricing helps detect and deter collusion or price manipulation. It makes the price signals provided by the market reliable and useful for participants in the market itself, in related markets and throughout the economy. Thus, liquidity, informational efficiency and market integrity go hand-in-hand and are mutually reinforcing.

3. Central Auction Markets: Some Basic Principles

The essence of open and competitive markets is the concentration of information. Once the different bits and pieces of information generated by multiple market participants are brought together, they are used to effect mutually beneficial trades. The nexus for the interaction of orders that generates trades is the trading pit (or floor) in the case of an open-outcry system or a central order book in the case of an electronic exchange. Order interaction is best accomplished in a central auction market which brings together all order information. As we discuss below, the concentration of the order flow is related to two phenomena that were studied in the market microstructure literature: liquidity externalities and mixing effects.

---

[58] As discussed in Section B.3(c) *infra*, evidence shows that this problem is particularly severe when smaller market participants reduce their market participation.

(a) Liquidity Externality

One of the key reasons for the development of central auction markets is the liquidity externality.[59] When a trader brings an order to the floor of the exchange (or to an electronic limit-order book), she does this to benefit herself. However, bringing the order to the exchange floor also improves the liquidity of the market as a whole: traders on the *other* side of the market have the option to trade against the order, thus making themselves better off. Therefore, an unintended by-product of the self-motivated placement of orders in the market is the creation of additional liquidity and depth.

Mendelson[60] showed that when more traders place orders in the market, this creates gains for *all* market participants over and above those perceived by the individual traders who placed the orders: "When traders join the market, they certainly gain some surplus as a result; but they also increase the surplus of all other traders by increasing the liquidity of the market (by providing traders on the other side of the market the opportunity to trade with them)."[61] That is, while the trader certainly benefits from the execution of his own order, the order creates additional benefits to the rest to the market, which the trader has no incentive to take into account.

Thus, market participants who bring orders to the exchange are not fully compensated for providing liquidity and immediacy to the rest of the market. As in any economic activity where the provider of a service is not fully rewarded for the benefit he confers on others, the level of service provided will be lower than is optimal for society as a whole. This implies that absent additional incentives or prohibitions, too little liquidity would be provided in trading markets. In this

[59] An externality is the effect of an economic action which is not fully priced in the market on the welfare of other economic agents. A positive externality occurs when an action generates an economic benefit to others without being fully rewarded for it by the market mechanism. It is well known that whenever such externalities exist, an intervention in market forces is called for to align individual incentives with the public interest; otherwise, the supply of the good or service in question will be sub-optimal.

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

situation, it should be the role of those responsible for overseeing trading to recognize the need for rules that increase the supply of liquidity beyond the level implied by individual incentives alone.

Trading markets are most competitive when *all* orders compete in the trading pit (or floor) or in an electronic trading forum. However, traders would sometimes prefer not to participate in the competitive process if they are allowed to trade in a non-competitive fashion. Even though the counterparties to the transaction could individually benefit from a non-competitive trade, the market as a whole would suffer. The liquidity externality means that *the benefits of competitive trading to the market as a whole exceed the benefits to the individual traders.* Thus, the Commission's requirement that with few exceptions, "All purchases and sales of any commodity for future delivery, and of any commodity option, on or subject to the rules of a contract market shall be executed openly and competitively by open outcry or posting of bids and offers or by other equally open and competitive methods"[62] is well-grounded in economic theory and empirical evidence.

The very existence of the liquidity externality—which implies that traders' choices must be constrained in order to bring about an overall optimum—is a sound reason for regulating trading markets in the first place. As we have written earlier in the context of securities markets, "Economic agents may take self-serving actions that reduce overall market liquidity. Public policies may thus have to complement private liquidity-enhancing actions. In fact, one of the key roles of the Securities and Exchange Commission is to facilitate and provide incentives to increase market liquidity, and to discourage actions that reduce liquidity."[63]

In the case of block trading, counterparties to the transaction would often prefer to consummate it to the exclusion of the rest of the market. These counterparties may argue that they

---

[61] *Ibid.*, at 271.
[62] 17 CFR § 1.38(a).

are sophisticated traders, they have arranged the trade using their own resources, they understand the options available in the rest of the market and they still prefer to trade just among themselves. However, the liquidity externality is one of the reasons for requiring that the block transaction still be integrated in some fashion with the rest of the market, since more trading interest at the exchange increases the liquidity of the entire auction market.[64] The liquidity externality creates a strong policy reason for looking beyond the choices made by individual traders on a trade-by-trade basis and examining their overall impact on the rest of the market.

(b) Mixing Effects

An auction market mixes different types of traders with different information in a way that makes it difficult to distinguish one type from another. This is necessary for the viability of the market. If the flow of traders is split between trading arenas in a way that enables market-makers[65] or trading counterparties to distinguish between them based on the information content of their trades, the quality of the market will suffer, and there may even be a market failure.

Market-microstructure theory distinguishes between two types of traders: "informed" traders whose trades are based on specific information about the traded instrument, and "uninformed" traders who trade for other reasons (e.g., when a seller needs cash). Researchers studying price formation in trading markets—Bagehot,[66] Black,[67] Glosten and Milgrom,[68] Kyle[69]

---

[63] Y. Amihud and H. Mendelson, "Liquidity, Asset Prices and Financial Policy," *Op. Cit.*, at 56-66.
[64] This is only one effect. Others will be discussed in Sections C and D, *infra*.
[65] Throughout, we use the term "market-makers" to designate traders who provide liquidity by quoting bids and offers in the market. Thus, market-makers may be locals in a trading pit, agents who place limit orders in an electronic system, or dealers who quote bid and ask prices.
[66] W. Bagehot, "The Only Game in Town," *Financial Analysts Journal* 27, March-April 1971, 12-14.
[67] F. Black, "Noise," *Journal of Finance* 41, 1986, 529-543.
[68] L. Glosten and P. Milgrom, "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders," *Journal of Financial Economics* 13, 1985, 71-100.
[69] A.S. Kyle, "Continuous Auctions and Insider Trading." *Econometrica* 53, 1985, 1315-1335.

and Admati and Pfleiderer,[70] among others—demonstrated that market liquidity is enhanced when uninformed traders participate in the trading process *together* with the informed traders, so the two types of traders are mixed and cannot be distinguished. The reason is that market-makers are reluctant to trade with counterparties who possess superior information because of the risk involved. For example, if an informed trader wants to sell, this means that the current market price is too high. Then, a market-maker who buys the instrument, hoping to sell it later and earn the bid-ask spread, is exposed to a higher risk that the instrument's price will decline and he will lose on the transaction. The more likely it is that the seller is "informed," the lower the bid price the market-maker is willing to pay (in order to protect himself). Analogously, if a prospective seller suspects that the buyer is informed, the ask (offer) price will be higher. The result is that in a market where transactions are more likely to be initiated by "informed" traders, the bid-ask spread is wider, market depth is smaller, market-impact costs are larger, and liquidity is lower. Mixing into the order flow "uninformed" liquidity-motivated trades helps reduce bid-ask spreads and market-impact costs and enhances liquidity.

Therefore, an additional benefit of a central auction market is that it *mixes* small and large, informed and uninformed trades. Indeed, smaller trades contribute to the performance of the market overall. Merton[71] showed that a larger investor base in an asset improves risk-sharing and diversification and lowers the equilibrium expected return (or yield) required by investors, making the asset more valuable. Evidence provided by Amihud, Mendelson and Uno[72] further supports this conclusion. They showed that for Japanese stocks whose institutional features were changed to

[70] A. Admati and P. Pfleiderer, "A Theory of Intraday Patterns: Volume and Price Variability," *Review of Financial Studies* 1, 1988, 3-40.
[71] This idea was presented when Nobel Prize Laureate Robert Merton was appointed president of the American Finance Association. See R. C. Merton, "A Simple Model of Capital Market Equilibrium with Incomplete Information," *Journal of Finance* 42, 1987, 483-511.
[72] Y. Amihud, H. Mendelson and J. Uno, "Number of Shareholders and Stock Values: Evidence from Japan," *Journal of Finance* 1998 (forthcoming).

facilitate the access of small individual investors to the market, the investor base increased substantially, leading to the appreciation of the affected stocks. Importantly, the increase in the number of individual shareholders was associated with an increase in the liquidity of these stocks.

Thus, a market designed in a way that excludes or discourages the participation of smaller trades reduces the trader base and hurts market liquidity, in addition to negatively impacting market integrity.[73] When more than one trading venue is available, and one of them is designed to facilitate the trading of large blocks at the expense of smaller trades, overall market liquidity suffers. And when the market is structured so that the order flow is split between two (or more) venues in a way that enables market-makers to distinguish between the different types of traders, market quality is reduced. Bessembinder and Kaufman[74] noted this problem when they found that segmentation of the order flow indeed hurts liquidity to the point where this market structure may be unsustainable: "Segmentation of the equity markets such that most price discovery takes place on the NYSE, while most liquidity trades are executed elsewhere, is unlikely to constitute a viable long-run equilibrium."[75]

In summary, the liquidity of a market can be improved by mixing together traders of different types—informed and uninformed, large and small—and treating them equally. Auction markets fulfill this role effectively.

4. Antecedents of Liquidity

In the foregoing analysis, we used liquidity, informational efficiency and integrity as evaluation criteria for the quality of a market. The positive effects of enhancing the latter two are

[73] See Section B.1(c), *supra*.
[74] H. Bessembinder and H. M. Kaufman, "A Cross-Exchange Comparison of Execution Costs and Information Flow for NYSE-Listed Stocks," *Journal of Financial Economics* 46, 1997, 293-319.

straightforward, and have already been discussed. In this Section we discuss the effects of enhancing liquidity.

The effect of liquidity on traders' welfare and on trading volume is clear: more liquid markets make traders better off and increase trading volume. A recent study by Wang, Yau and Baptiste[76] on the effect of illiquidity costs on trading in the futures markets examined the cross-sectional relationship between trading volume and transaction costs for seven futures contracts on the following underlying commodities: T-Bonds, S&P 500 index, Deutsche mark, wheat, soybeans, copper and gold. For *all* seven contracts they found that there is a strong negative relationship between transaction costs and trading volume. That is, a rise in transaction costs reduces trading activity (after controlling for other factors).

More subtle is the effect of liquidity on the values of traded assets. In securities markets, liquidity was found to have a strong positive effect on asset values because investors demand a compensation for the illiquidity costs they bear. Amihud and Mendelson[77] presented a theory showing that across securities, the lower the liquidity of a security, the lower its price. Although the illiquidity cost per transaction is usually small relative to the security's price, the cumulative effect of illiquidity costs on the security's *value* is considerable because they are incurred *repeatedly* whenever the security is traded.

The empirical evidence strongly supports this theory. Amihud and Mendelson[78] found that the average returns on NYSE stocks over the 1961-1980 period were higher for stocks with wider bid-ask spreads, after controlling for risk. If stocks have to generate higher returns to compensate for lower liquidity, their values are lower. Recent evidence for NASDAQ stocks strongly supports

---

[75] *Op. Cit.*, at 295.
[76] Wang, Yau and Baptiste, *Op. Cit.*
[77] Y. Amihud and H. Mendelson, "Asset Pricing and the Bid-Ask Spread," *Journal of Financial Economics* 17, 1986, 223-249.

the negative relationship between bid-ask spreads and stock prices.[79] Other studies use different measures of liquidity but arrive at the same finding: the lower the liquidity of stocks, the higher the required return by investors, or the lower their price.[80]

Higher liquidity also has a beneficial effect on bond prices; or, conversely, if a trading system provides lower liquidity on bonds, the yield to maturity required by investors rises. Amihud and Mendelson showed this empirically, finding that U.S. Treasury Securities with identical cash flows had different yields to maturity, depending on their liquidity: the less liquid debt securities had higher yields.[81] Specifically, they compared the yields on U.S. Treasury bills and notes with the same time-to-maturity of less than six months. The two financial instruments generate identical cash flows, but the bills are more liquid than notes. For example, the Treasury bills in the sample had an average bid-ask spread of 0.0078% of value, whereas the average bid-ask spread on the corresponding notes was 0.03% of value, about four times wider.[82] These differences in liquidity translated into substantial differences in value: the average annual yield on notes was 43 basis points higher than on bills with the same maturity. Thus, when a U.S. Treasury security is less liquid, investors require a higher yield to compensate for the increase in illiquidity costs.

---

[78] *Ibid.*

[79] V. Eleswarapu, "Cost of Transacting and Expected Returns in the NASDAQ Market," *Journal of Finance* 52, 1997, 2113-2127.

[80] See M. J. Brennan and A. Subrahmanyam, "Market Microstructure and Asset Pricing: On the Compensation for Market Illiquidity in Stock Returns," *Journal of Financial Economics* 41, 1996, 441-464. They use the market impact as a measure of illiquidity costs. See also V. Datar, N. Naik and R. Radcliffe, "Role of Trading Activity in the Cross-Section of Stock Returns," Working Paper, London Business School, 1995. They use as a measure liquidity the stock turnover (ratio of trading volume to number of shares outstanding) and find a very strong and statistically significant relationship: the higher the liquidity (thus measured), the lower the average return earned on stocks, after controlling for risk.

[81] Y. Amihud and H. Mendelson, "Liquidity, Maturity and the Yields on U.S. Treasury Securities," *Journal of Finance* 46, 1991, 1411-1425. Confirming evidence is presented in A. Kamara, "Liquidity and Short-Term Treasury Yields," *Journal of Financial and Quantitative Analysis* 29, 1994, 403-417. The relationship between yield differentials and liquidity differentials was first noted in K. Garbade, "Analyzing the Structure of Treasury Yields: Duration, Coupon and Liquidity Effects," in *Topics in Money and Securities Markets*, BANKERS TRUST CO., 1984.

[82] The bid-ask spread relative to value was calculated as (ask price - bid price)/(ask price + accrued interest), where the accrued interest on bills was zero.

Futures trading contributes to an increase in the liquidity of the underlying security. Erwin and Miller[83] showed a significant decrease in the bid-ask spread upon addition of a stock to the S&P 500 list, on which there are futures contracts traded. This benefit accrued only to stocks that were not trading listed options, suggesting that exchange trading of a derivative instrument increases the liquidity of the underlying. Consistent with the liquidity theory, Erwin and Miller found that "the decrease in the bid/ask spread for nonoptioned stocks is accompanied by a significant and permanent increase in share price and trading volume… The price and volume effects associated with S&P addition reflect enhanced stock liquidity. The decrease in the bid/ask spread for nonoptioned stocks is attributed to informational efficiencies achieved via index arbitrage trading…"[84]

The above studies show that the impact of liquidity on asset values is sizable. Existing studies focusing on the case of equities show that consistent with the theory, improved liquidity leads to an increase in securities values. Indeed, Amihud, Mendelson and Lauterbach[85] showed that when the trading mechanism in a stock market was improved to provide better liquidity, stock prices rose by 5 to 6%. Conversely, subjecting a stock to a trading mechanism that reduces its liquidity reduces its value.[86]

Increasing the liquidity of a financial instrument therefore increases its value or reduces the yield required by investors. Further, a liquid futures market improves the quality of the market for the underlying security due to the beneficial effect it has on the cost of hedging, informational efficiency and trading volume.[87] It follows that the design of the trading mechanism for futures

[83] G. R. Erwin and J. M. Miller, "The Liquidity Effects Associated with Addition of a Stock to the S&P 500 Index: Evidence from Bid/Ask Spreads," *The Financial Review* 33, 1998, 131-146.
[84] *Ibid.* at 131.
[85] Y. Amihud, H. Mendelson and B. Lauterbach, *Op. Cit.*
[86] Evidence from the Tel Aviv Stock Exchange shows that transferring a stock to a less-liquid trading mechanism reduced its value.
[87] See Section B.1(b), *supra*.

contracts on U.S. Treasury Securities affects the yields on the underlying securities. A well-designed trading mechanism for these futures contracts will increase the liquidity of the underlying Treasury Securities and will thus reduce the yield required by investors, translating into a lower cost of debt for the U.S. Government. Conversely, impairing the liquidity of the futures market will have a negative effect on the underlying Treasury Securities, as a less liquid futures market increases the costs of hedging, leading to higher required yields. Intuitively, the futures markets can serve as an "insurance policy" against changes in the prices of U.S. Treasury Securities. When the cost of the "insurance policy" increases, risk-averse investors will require a higher yield on Treasury Securities, raising the U.S. Government's cost of debt.

5. Futures Market Structures: Electronic Exchanges vs. Open Outcry

In recent years, there has been an increase in the use of electronic futures exchanges, which have become viable due to developments in information technology, globalization and the experience gleaned from the successes and failures of electronic contenders in the past. In April of this year, the French exchange, MATIF, adopted the Paris Bourse's NSC-VF Platform alongside its traditional open outcry system. Six weeks later, it abandoned its open outcry system altogether. Also this year, London's LIFFE lost its lead in trading the German Bund futures contract to Germany's DTB, an electronic exchange.

In an electronic exchange, both the communication system and the execution system are computer-based. That is, traders enter orders into a computer terminal, personal computer or workstation and the orders are communicated to a central system that concentrates the order flow. The system matches buy and sell orders automatically, following the priority rules that were defined as part of the system's design (typically based on price and time priority), executes the trades and creates electronic clearing records.

The choice between open outcry and electronic exchanges is not clear-cut: there is a tradeoff between the two approaches. However, the cost trends clearly favor electronic exchanges: whereas the costs associated with physical trading facilities have not declined over time, the costs of electronic-based systems have been decreasing at a rapid rate. This means that in the very long term, electronic exchanges are likely to prevail. Yet, it is possible that open outcry systems will remain competitive for many years to come, and, although direct comparisons of illiquidity costs on markets that operate simultaneously are scarce, in a number of cases the open outcry market is more liquid. For example, Vila and Sandman[88] compared bid-ask spreads in two markets that trade futures contracts on the Nikkei Stock Average in overlapping hours: The Singapore International Monetary Exchange (SIMEX), an open outcry market, and the Osaka Securities Exchange (OSE), a computer-based market. They found that the bid-ask spread on SIMEX is lower than the bid-ask spread in OSE, showing that an open outcry market may in fact be more liquid than an electronic exchange. We briefly summarize the tradeoffs between the two types of systems below.

(a) Advantages of Electronic Exchanges

The primary advantage of automated systems is operating cost. The cost of information technology is decreasing at a phenomenal rate, following "Moore's Law": the number of logic elements on a chip doubles every eighteen months, and the cost of information processing correspondingly declines. Human labor cost, on the other hand, has been steadily increasing over time. Given these cost trends, substituting electronics for human labor can greatly reduce the cost of operations.

---

[88] A.F. Vila and G. Sandmann, "Floor Trading versus Electronic Screen Trading: An Empirical Analysis of Market Liquidity and Information Transmission in the Nikkei Stock Index Futures Market," Working Paper, London School of Economics Financial Markets Group, October 1995.

The second advantage of electronic exchanges is speed. Whereas open outcry systems are limited by the speed of human processing capabilities, electronic systems are limited only by the speed of light. Related to the speed factor is processing capacity: open-outcry systems are limited by the processing capacity of individuals and by the physical size of the trading pit. Expanding the capacity of open-outcry systems thus requires at best hiring and at worst rebuilding the physical space of the exchange. In contrast, properly designed electronic exchanges may be easily expandable: increasing capacity can be as straightforward as purchasing a more powerful computer.[89]

Another potential advantage of electronic exchanges is technological innovation. The ability to process large amounts of information at high speed creates the opportunity to restructure the trading process and take advantage of the new capabilities. Anonymity, for instance, can be more easily facilitated in electronic exchanges. More importantly, electronic systems permit the implementation of smart algorithms to facilitate different types of transactions. This means that to realize their full potential, electronic exchanges should go beyond just computerizing the floor-based trading process and actually take advantage of the capabilities of new technology.

Neither of these advantages are costless, however. Achieving the potential advantages of new technology requires exchanges to stay on the leading edge of technological advances and to make sizable investments in software development, data-communications infrastructure, storage and information processing resources. And, being truly innovative requires the development of custom software, which entails substantial development, testing and maintenance costs, as well as significant risks.

[89] Although in practice, this is rarely the case.

(b) Advantages of Open-Outcry Systems

The first advantage of open-outcry systems is due to the richness of human interaction. Human communication permits many forms of interaction, combining hand signals, voice, facial expression and "body language" that are virtually impossible to replicate on a computer screen. With current computer technology, a great deal of richness is lost in the transition to electronic media: the atmosphere in the pit is difficult to replicate on a computer screen. While it is likely that information technology will permit richer forms of communication in 10-20 years, the design of today's electronic exchanges does not take advantage of such capabilities.

A second advantage of the open-outcry system is the flexibility inherent in any human-based system. While the cost of running an electronic exchange may be lower than the cost of operating a trading pit, it requires a substantial fixed investment in systems development. Any changes in the operating rules (or further yet, in the logic) of computer-based systems require costly and risky system upgrades, and going through a painful conversion process.

Electronic exchanges cannot easily cope with unanticipated circumstances that were not contemplated in the system's design: the computer-based systems are simply not flexible enough to allow for unusual or unforeseen circumstances. A human-based open-outcry system, on the other hand, can accommodate unanticipated events more gracefully: for example, a human decision-maker may allow temporary changes in order to adapt to unforeseen circumstances. It is the unusual and unanticipated occurrences that really test the integrity of the system, and under these conditions open-outcry systems have an advantage.[90]

---

[90] Indeed, Franke and Hess found that in times of stress, order flow migrated to the open outcry exchange when one was available. See G. Franke and D. Hess, "Information Diffusion in Electronic and Floor Trading," Working Paper, University of Konstanz, September 1997.

Related to flexibility is another advantage of open-outcry systems: lower risk. Open-outcry has been around for many years, and has thus been tested under a variety of circumstances. In contrast, electronic systems are still subject to a variety of malfunctions, increasing the risk of operations. And, because human-based operations are more flexible, open outcry systems are adaptable in the event of crisis. Electronic systems are more risky in that respect, and may have to be shut down in the event of malfunction.

Finally, open outcry enables traders to handle complex, interdependent orders that are not supported by most electronic exchanges. Futures trades often involve the implementation of complex strategies across different contracts, which can often be accomplished faster and more easily in the trading pits. This is not to say that automated systems are inherently incapable of dealing with such situations: to the contrary, we have repeatedly proposed – starting many years ago—that such innovative systems would reap the true benefits of automation.[91] However, most of today's electronic exchanges fall short of meeting this challenge, and this approach is still largely untested.

(c) An Approach to Automation

In our earlier work,[92] we proposed an approach to the development of computer-based trading systems that was based on three fundamental principles:

(1) Competition between alternative trading mechanisms;

(2) Flexibility in operation; and

---

[91] See, e.g., Y. Amihud and H. Mendelson, "An Integrated Computerized Trading System," in *Market Making and the Changing Structure of the Securities Industry*, Lexington Heath, 1985, 217-235; Y. Amihud and H. Mendelson, "The Effects of Computer-Based Trading on Volatility and Liquidity," in *The Challenge of Information Technology for the Securities Markets: Liquidity, Volatility, and Global Trading*, Dow Jones-Irwin, 1989, 59-85.
[92] *Ibid.*

(3) Integration (through appropriate interfaces) of the three phases of the trading process discussed in Section B.2, *supra.*[93]

We proposed that the key dimension of competition is competition between *trading mechanisms*, not necessarily between exchanges. When competing trading mechanisms operate within a given exchange, market participants can choose between them while still preserving the possibility of price improvement across mechanisms. In the case of open outcry operating side-by-side with an electronic exchange, market participants can have access to both, and the exchange can increase flexibility and facilitate the choice of a trading mechanism across the competing alternatives.

In this way, exchanges can try new approaches and let the market decide which ones work best. This approach enables markets to try out new trading systems, using them on an experimental basis for a subset of contracts, and extending their use if they are successful. This approach also serves to reduce the risk inherent in the development of new computer-based systems. It enables exchanges to first implement a "beta" version of the system, to correct any remaining design flaws and to hone the system's parameters based on the experience gained from limited use. Once experience shows the system is working smoothly, it can be implemented on a full-scale basis.

[93] These phases are information assimilation, order communications and execution.

C. MARKET FRAGMENTATION AND INTEGRATION: ISSUES AND RESEARCH FINDINGS

The introduction of a new market with a new trading system, call it **B**, to trade an instrument that is already traded in an existing market, say **A**, has two opposing effects. Seemingly, the introduction of **B** is beneficial because it promotes competition between the two markets, yielding the benefits that are commonly associated with competition, such as reduction in execution costs[94] and technological innovation. However, trading the same instrument in two independent markets causes *market fragmentation*.

We defined a trading market as a mechanism that serves to communicate orders for trading an instrument and to transform these orders into trades.[95] When this transformation is informationally centralized, i.e., all the order data are available when orders are transformed into trades, the market is *consolidated*: all bids interact with all offers, yielding the optimal market-wide execution.[96] The market is *fragmented* when orders are decomposed into distinct subsets that do not fully interact, such as trading the same instrument in two independent markets. As a result, potential mutually beneficial trades are missed and the quality of execution is inferior to that obtained in a consolidated market. Yet, as discussed below, fragmentation may result from the desires of some traders to trade in a separate market.

Given the advantages of a central auction market,[97] other things being equal, consolidation of the order flow into a single market would lead to a superior market structure. It follows that there is a tradeoff between the costs of fragmentation and the benefits of intermarket competition.

---

[94] See Y. Amihud and H. Mendelson, "A New Approach to the Regulation of Trading Across Securities Markets," *Op. Cit.*, for a survey of the tradeoff involved.

[95] See Section B.2, *supra*.

[96] See, H. Mendelson, "Consolidation, Fragmentation and Market Performance," *Journal of Financial and Quantitative Analysis* 1987, 189-207. A consolidated market is often called a *central auction* market.

[97] See Section B.3, *supra*.

While competition is beneficial, market fragmentation has negative effects on market liquidity, on the price discovery process and on market integrity. The question is what is the net effect, or, put differently, whether the benefits of competition are sufficient to offset the negative effects of fragmentation.

We evaluate the findings of past and current research on this question along the three key dimensions of market quality outlined in Section B.2: liquidity, informational efficiency and market integrity.

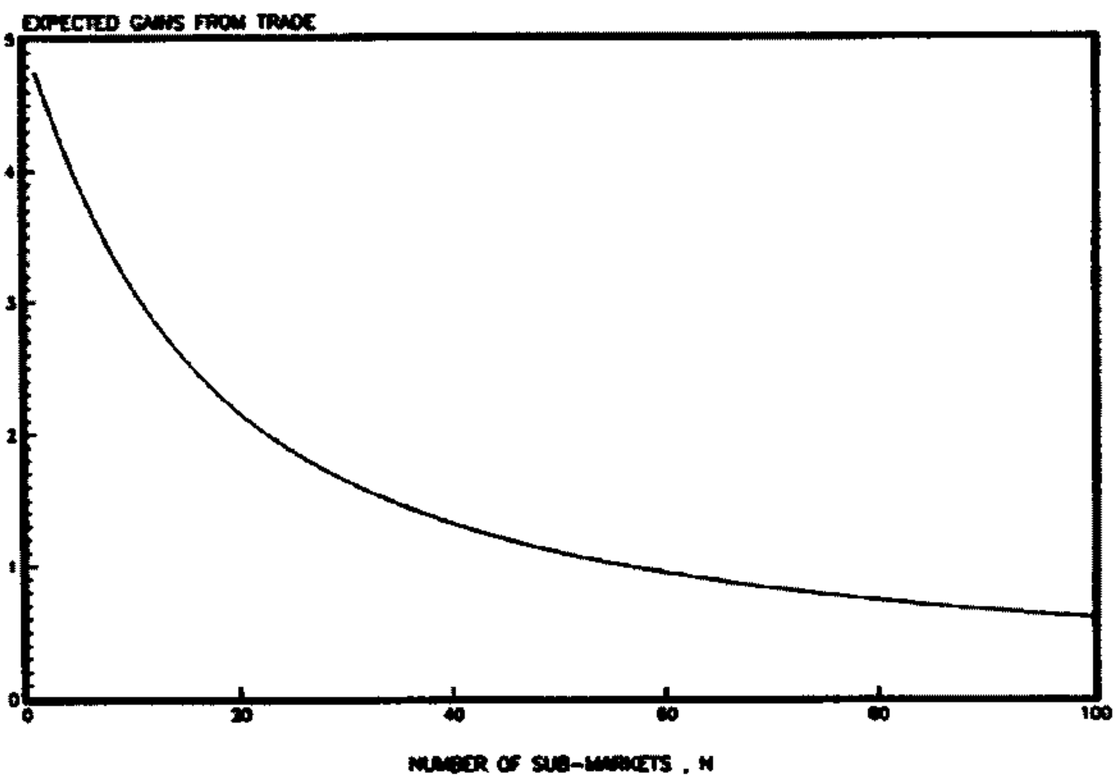
1. Fragmentation and Market Liquidity

The key dimension of competition in a system of trading markets is competition among *market participants* rather than among trading markets. In a central auction market, competition among market participants is the cornerstone of the market's operating rules. An auction market that enables any market participant to compete with any other market participant has in fact built-in *internal competition*. This *intra*market competition disciplines all market participants and brings about the advantages of liquidity, efficiency and integrity discussed in Section B.2, *supra*. Mendelson[98] compared the performance of a number of consolidated and fragmented market architectures. His performance measures are the overall gains from trade,[99] the price volatility faced by market participants, the quantity traded and the "noise" (random disturbances) in the reported transaction prices. Comparing a consolidated order flow sent to a central auction market to an order flow that is split between fragmented markets, Mendelson found that fragmentation increases the volatility of transaction prices and reduces the overall gains from trade as well as the

[98] *Ibid.*

quantity traded. Mendelson used the overall gains from trade as a measure of market liquidity. His Figure 4, replicated here as Figure C-1, shows that this measure of liquidity declines as the market becomes more fragmented. These results are directly related to the liquidity externality discussed in Section B.3(a): when markets are fragmented, each order is exposed only to a fraction of the order flow, and hence enjoys less liquidity, because the order flow in each market fragment is smaller than it is in the consolidated market. Hence, the likelihood that a trader can find a counterparty with whom to trade is correspondingly smaller.

Figure C-1: Liquidity declines as the market becomes more fragmented.



[99] The overall gains from trade are obtained by adding up, over all traders whose trades were executed in the market, the benefits they gained from trading compared to the alternative of not trading. Thus, the overall gains from trade measure the value of having in operation the market trading the instrument.

Cohen, Maier, Schwartz and Whitcomb[100] studied the effect of having, in addition to the main market, order execution in "satellite" markets managed by brokers off the exchange floor. They showed that while this is beneficial for the brokers, it harms the market as a whole. While some traders might do better for themselves in a fragmented environment, fragmentation reduces overall welfare because the fragmented market leads to a wider bid-ask spread and greater price uncertainty.[101] This is another demonstration of the externality effects associated with the exchange process. Cohen, Maier, Schwartz, and Whitcomb thus concluded that the best market structure is an informationally-consolidated market. Along similar lines, Cohen, Conroy and Maier[102] showed that in a fragmented market, there is a lower chance that an order will find an available counterparty, the expected time a limit order has to wait until it is executed is longer, and the bid-ask spread is wider. Thus, fragmented markets reduce liquidity.

Khan and Baker[103] provided evidence on the negative effect of fragmentation on liquidity by examining the case where regional exchanges attracted trades in listed stocks away from the main exchanges (NYSE and AMEX). On the one hand, competition from the regional exchanges should improve liquidity, but on the other hand, fragmentation of the order flow should hurt liquidity. The evidence showed that only low-liquidity stocks benefited from the competition provided by the regional exchanges. The trading in these stocks is usually dominated by the NYSE specialist and there is very little *intra*market competition in their trading. Thus, the competition introduced by the regional exchanges had a net beneficial effect. However, the liquidity of large, actively traded stocks was hurt by the introduction of competition from the regional markets. These stocks apparently enjoyed a great deal of *intra*market competition that made them quite

[100] K.J. Cohen, 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* 6, 1982, 117-136.
[101] That is, prices are less informative: the variance of price around the true value of the instrument is larger.
[102] K.J. Cohen, R.M. Conroy and S.F. Maier, *Op. Cit.*

liquid on the main exchange, and thus the added competition from the regional markets could add very little benefit. On the other hand, the diversion of part of the order flow from the NYSE and AMEX to the regional exchanges caused market fragmentation. On balance, the negative effects of fragmentation more than offset the benefits of competition. Kahn and Baker's evidence showed that the average effective bid-ask spread of the liquid stocks in their sample increased by over 50% as a result of fragmentation—from 0.92% before to 1.39% after. Large stocks with an increase in the bid-ask spread outnumbered large stocks with a decrease in the bid-ask spread by a ratio of 2:1.

These results indicate that fragmentation of the order flow for actively traded instruments on an exchange into independent order flows in separate exchanges hurts liquidity when the central marketplace was sufficiently competitive. While trading in multiple markets may increase intermarket competition, it does not add much to the competition for order flow when there is already a great deal of competition between market participants. Yet, the resulting fragmentation has negative effects that outweigh the smaller benefits of competition. These results strongly indicate that in futures markets with a great deal of internal competition, allowing the diversion of order flow to another market is likely to hurt liquidity.

Dual listing of the same stock in two segmented markets was also shown to have a negative effect on stock prices. Fang[104] examined the price effect of listing Chinese stocks traded on the Shanghai Stock Exchange for dual trading on the Hong Kong Stock Exchange ("HKSE"). Such listing clearly added a dimension of competition to the market, having the established HKSE compete with the Shanghai Stock Exchange. However, Fang pointed out that the two markets are segmented. He compared the pre-listing prices of the dually-listed stocks to their prices 40 days

[103] W. Khan and H. Baker, "Unlisted Trading Privileges, Liquidity and Stock Returns," *Journal of Financial Research* 16, 1993, 221-236.
[104] Z. Fang, "Dual Listing Impacts on Chinese Stock Returns: Shanghai and Hong Kong," Working Paper, City University of Hong Kong, 1997.

following the dual listing on the HKSE (controlling for the market effect). The prices of the dually-listed stocks declined as a result of the dual listing. This, again, suggests a negative effect of fragmentation.

Evidence from the U.S. also shows that market fragmentation leads to lower-quality execution in "satellite" markets. Lee[105] analyzed the execution prices of similar adjacent trades of stocks both listed on the NYSE and also traded off-board, in regional exchanges and on the NASDAQ Stock Market. His results show that the central auction market provides the best execution prices, because traders in the auction market can benefit from price improvements that ensue from the competition there. Lee found that "the average price differences between the NYSE and matched off-board trades is 0.7 to 1 cent per share. These results are statistically significant…" (p. 1035), and concluded that this resulted from market fragmentation. Again, competition between traders on the NYSE floor is more effective than intermarket competition: the former led to price improvement whereas the latter actually worsened prices.

Further empirical evidence on the net effect of intermarket competition versus fragmentation is provided by market developments following the revisions in NYSE Rule 390 that became effective on July 18, 1980. NYSE Rule 390 forced consolidation of the order flow in NYSE-listed stock by restricting NYSE member firms from executing orders to trade these stocks off the exchange floor. The SEC enacted Rule 19c-3 which, as of July 1980, enabled fragmentation of the order flow by allowing member firms to trade off the floor stocks that were listed after April 26, 1979. Effectively, Rule 19c-3 allowed exchange members to make off-Board markets. This fact situation enables an examination of the net effect of intermarket competition versus fragmentation by comparing the execution costs on NYSE-listed stocks that were subject to Rule

---

[105] C. M. C. Lee, "Market Integration and Price Execution for NYSE-Listed Securities," *Journal of Finance* 48, 1993, 1009-1038.

390 to the execution costs for stocks that were covered by Rule 19c-3. The SEC assumed that allowing off-board trading would foster competition between the NYSE and alternative markets. Yet, the new rule fragmented the order flow.

Bessembinder and Kaufman[106] studied whether the benefits of competition created by Rule 19c-3 were sufficient to offset the negative effects of fragmentation that the Rule caused. They compared actual transaction costs for 300 NYSE-listed stocks that were also traded on regional exchanges and on the NASDAQ stock market. The sample of 300 stocks was divided about evenly between 19c-3 stocks and non-19c-3 stocks. Bessembinder and Kaufman measured transaction costs by examining the *effective* bid-ask spread, which measures the (absolute) difference between transaction prices and the midprice between the quoted bid and ask prices.[107] They found that execution costs were substantially larger for 19c-3 stocks. These results were statistically significant, and remained unchanged after accounting for the different characteristics of 19c-3 and non-19c-3 stocks.

Bessembinder and Kaufman[108] also measured transaction costs by the transitory "market impact cost." This measure of execution cost was also generally larger for 19c-3 stocks. They concluded that execution costs were higher for securities that were eligible for Rule 19c-3. These results suggest that the costs of market fragmentation outweigh the benefits of competition. Bessembinder and Kaufman conclude (p. 316): "In contrast to the apparent intent of Rule 19c-3, these results provide no support for the rationale that allowing NYSE member firms to execute trades off the NYSE effectively increases competition or reduces trade execution costs." Similar

---

[106] H. Bessembinder and H. M. Kaufman, " *Op. Cit.*, at 293-319.
[107] This is designed to take into account the fact that orders may be executed in-between the quoted bid and ask prices.
[108] Bessembinder and Kaufman, *Op. Cit.*

results were obtained by Davis and Lightfoot.[109] Indeed, the competition within the NYSE auction market was sufficient, and, in addition, the order flow was reasonably consolidated by Rule 390, which prohibited off-Board trading by member firms. Given the competition that already existed on the floor of the NYSE, Rule 19c-3 did not add much in terms of the benefits of competition, yet it caused harm because of fragmentation. For the securities that were not subject to Rule 19c-3, the consolidation of the order flow was beneficial.

Another cost of fragmentation is the search and delay cost discussed in Section B.2(a), *supra*. When the order flow is split between two or more markets, the cost of searching for the best price is greater, because prices have to be compared in two separate locations. While looking for a better price may lead to a better result in executing the order, in fast-moving, volatile futures markets, this entails a search and delay cost to traders.

2. Fragmentation and Informational Efficiency

In a fragmented market, where the same instrument is traded in (at least) two locations, a smaller market can benefit from the information produced in the main market and free-ride on it. This leads to insufficient production of information, which in turn negatively affects the price discovery process.

Garbade and Silber[110] examined intermarket information transmission for stocks that were multiply-traded on the NYSE, the Midwest Stock Exchange (now the Chicago Stock Exchange) and the Pacific Stock Exchange. They found that the regional exchanges are best characterized as

---

[109] J.L. Davis and L. E. Lightfoot, "Fragmentation versus Consolidation of Securities Trading: Evidence from the Operations of Rule 19c-3," *Journal of Law and Economics* 41, 1998, 209-238.
[110] K. Garbade and W. Silber, "Dominant and Satellite Markets: A Study of Dually-Traded Securities," *Review of Economics and Statistics* 61, 1979, 455-460.

satellites of the NYSE, meaning that the prices set on regional exchanges are essentially derived from NYSE prices with some time delay allowing for price adjustment.

Hasbrouck[111] provided more recent evidence of this effect. He analyzed the quoted prices of stocks traded on the NYSE and the best quotes of the same stocks that are traded on regional exchanges. Hasbrouck concluded (p. 1197): "The results suggest that price discovery appears to be concentrated at the NYSE: the median information share is 92.7 percent." This means that the regional exchanges free-ride on the information produced in the main market, the NYSE, and their contribution to the price discovery process is very small. The share of the NYSE in trading volume was 85%, smaller than its share in the price discovery process. That is, the share of trading that the satellite markets attract away from the main market, the NYSE, is far larger than the contribution of these markets to price discovery. These results imply that the satellite exchanges gain at the expense of the main market, the NYSE: while the NYSE bears the costs of price discovery, the satellite exchanges free-ride on NYSE's prices and manage to get a larger share of trading volume.

Similar results are obtained by Subrahmanyam,[112] who investigated the informativeness of trades on the NYSE and AMEX vis a vis the regional exchanges and the OTC market, using intraday transaction data for a sample of 1,258 stocks. He found that for a significant proportion of stocks, the transactions in the main exchange (NYSE or AMEX) had a stronger information effect on prices than transactions on secondary exchanges. That is, the satellite markets feed on information provided by the main market, where the price discovery takes place, without contributing much to the process.

[111] J. Hasbrouck, "One Security, Many Markets: Determining the Contribution to Price Discovery," *Journal of Finance* 50, 1995, 1175-1199.

[112] A. Subrahmanyam, "Multi-Market Trading and the Informativeness of Stock Trades: An Empirical Intraday Study," *Journal of Economics and Business 49, 1997, 515-531.*

Bessembinder and Kaufman[113] examine the information effect on prices of transactions of stocks that are traded both on the NYSE and on satellite markets – the regional exchanges and the NASD dealer market. They found that the information effect on prices is nearly twice as high on the NYSE than it is in the satellite markets, especially compared to the NASD market. They concluded that the alternative markets diverted uninformed order flow from the NYSE, resulting in "cream-skimming": the satellite markets took very little risk in providing market-making services, since their order flow came from small retail investors. They could free-ride at low cost on the information and liquidity that were provided by the NYSE, where the more risky executions took place. Bessembinder and Kaufman suggest that the segmentation of the market in NYSE-listed stocks, whereby most price discovery takes place on the NYSE while liquidity-motivated trades are executed elsewhere, operates to the detriment of the markets in the long run.

An extreme form of free riding is provided by *crossing networks,* which do not even attempt to provide a price discovery function.[114] Examples of crossing networks in the equity markets are Reuters' Instinet Crossing Network and ITG's POSIT. While the traders choosing to trade on these networks clearly find this to their benefit, the overall effect on the market is often negative due to fragmentation. Hendershott and Mendelson[115] studied the effects of this form of fragmentation on existing dealer markets.[116] They showed that when traders strategically use the established market as a "market of last resort," the main market becomes more risky, resulting in a widening of the bid-ask spread. This, in turn, reduces the attractiveness of the main market. Thus, even when crossing networks offer effective, low-cost execution services, if their use grows to the point where

[113] Bessembinder and Kaufman, *Op. Cit.*
[114] For a discussion of the proposed CFFE's version of a crossing network, called "Market Crossing," see Section E.6, *infra*.
[115] T. Hendershott and H. Mendelson, "Crossing Networks and Dealer Markets: Competition and Performance," Working Paper, Stanford University, 1998.
[116] While their analysis is couched in terms of a dealer market, similar results apply to other forms of market organization.

they can affect the main market, they can actually have an aggregate net adverse effect on traders. Consequently, the equilibrium behavior of individual traders does not give rise to an efficient outcome. In fact, the bid-ask spread in the main market can increase to the point where it cannot cover its costs and it has to shut down. Indeed, Bessembinder and Kaufman found that the problem of market fragmentation "is reinforced by the increasing use of electronic crossing networks such as Instinet and Posit which may also divert uninformed order flow from the NYSE" and may result in an unsustainable long-run equilibrium.[117] The evidence surveyed above indicates that the less informative trades take place in "satellite" markets and the more informative ones are concentrated in the main market.

Market liquidity crucially depends on the mixing of informed and uninformed traders in the same market (see Section B.3.(b), *supra*). The evidence cited above shows that satellite markets contribute to the separation of informed from uninformed traders by diverting uninformed traders from the main market (what Bessembinder and Kaufman call "cream skimming"). This may well reduce the liquidity of the main market, which is more informative, and consequently reduce the quality and integrity of the overall market. In contrast, requiring the two types of traders to mix in the same market will help overall liquidity. The deterioration in market efficiency, or the increase in price uncertainty, due to fragmentation is consistent with the theoretical findings in Mendelson[118] and Cohen, Maier, Schwatrz and Whitcomb[119] that were reviewed in the previous subsection. Figure C-2 replicates Mendelson's Figure 3, and shows that as the degree of market fragmentation increases, so does the uncertainty (measured by the price variance) faced by traders in the market.

[117] Bessembinder and Kaufman, *Op. Cit.*, at 295.
[118] Mendelson, *Op. Cit.*
[119] Cohen, Maier, Schwartz and Whitcomb, *Op. Cit.*

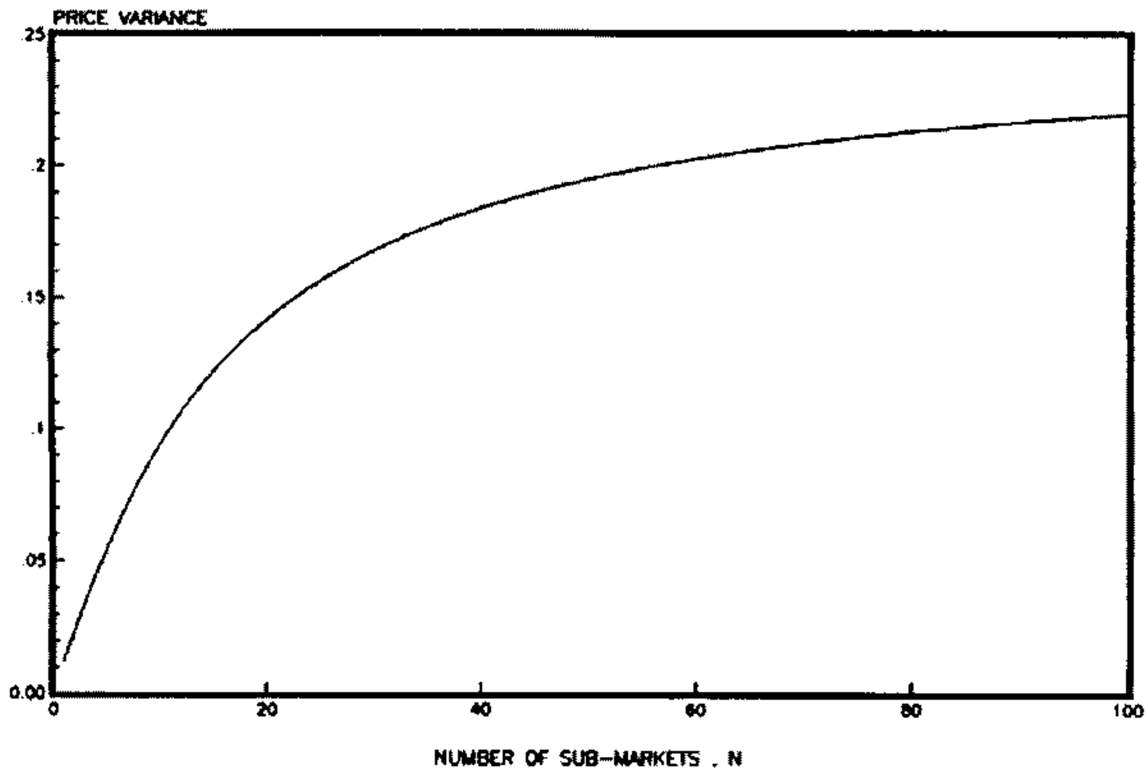


Figure C-2: Price variance increases as market fragmentation increases.

The fragmentation of the market by trader type – for example, having one market that serves large block trades and another that serves smaller traders – has other negative effects. If large investors with market power can negotiate better terms for executing their orders, retail investors have to pay more to cover the costs of market-makers who provide liquidity. This was illustrated in the London Stock Exchange, where the bid-ask spreads paid by retail traders were substantially greater than those for large institutional traders.[120]

[120] New York Stock Exchange, Research & Planning Division, "Auction Market Trading," October 13, 1992, at 9.

The effects of market fragmentation on informational efficiency and price discovery were analyzed by Madhavan.[121] As in other studies of informational efficiency, trading leads to price discovery, whereby the instrument's transaction prices approach its underlying value. The more efficient the market, the smaller the deviations of transaction prices from the underlying value. Madhavan found that in a fragmented market with an incomplete flow of information available to all market participants, price volatility is higher and price efficiency is impaired. In addition, Madhavan showed that the beneficiaries from the lack of transparency are the *large* and *better-informed* traders, who extract rents from their superior information through dynamic trading, at the expense of uninformed liquidity traders who join the trading process later on. Small liquidity traders, on the other hand, prefer a transparent market. Madhavan showed that large traders benefit from shifting their trading from a transparent to a less transparent market and thus fragment the order flow. Madhavan summarized the effect of market fragmentation on transparency as follows (p. 593): "Large traders essentially 'front run' their own trades, and nondisclosure allows these traders to obtain better execution overall. This natural demand for less transparent trading systems suggests that some degree of fragmentation may be inevitable unless there is mandatory trade disclosure."

The evidence is consistent with the theory. The London stock market, whose rules allowed delayed reporting of block trades and was thus less transparent, attracted large traders from other European markets with strict disclosure rules, e.g., from the Paris Bourse.[122] The block traders who shifted their trading to London personally benefited, but the markets that required transparency were circumvented. This shift in trading forced the markets into a "race to the bottom," where the

---

[121] A. Madhavan, "Consolidation, Fragmentation, and the Disclosure of Trading Information," *Review of Financial Studies* 8, 1995, 579-603.
[122] See A. Benos and M. Crouhy, "Changes in the Structure and Dynamics of European Securities Markets," *Financial Analysts Journal* 52, 1996, 37-50.

more transparent markets had to relax their disclosure requirements so as not to lose trading volume.

A similar "race to the bottom" ensues whenever traders have an incentive to divert order flow to markets that do not follow the strict rules of a central auction market. Whereas individual traders can benefit in the short term from such diversion, the performance of the market as a whole deteriorates. First, as discussed in Section C.1, market fragmentation harms liquidity and increases trading costs. Second, as discussed in this Section, fragmentation hurts the informational efficiency of the market and the price discovery process. A degradation in the quality of market prices affects traders, who are subject to price uncertainty, and decision makers throughout the economy who use the market prices to make tradeoffs, business and policy decisions. This also affects market integrity, which is discussed next.

3. Fragmentation and Market Integrity

The "law of one price" and the tenet of "best execution" are hallmarks of market integrity. However, both may be violated in a fragmented market: orders may be simultaneously executed in the market at different prices, and there may even be *trade-throughs*, namely, a situation where a trade is carried out in one market at a price that is worse (lower for sell orders and higher for buy orders) than the best quoted price available in another market.

This may seem odd, because it is apparently in a trader's own best interest to execute his order at the best available price.[123] To the extent this is consistent with investors' interests, one would believe that traders would search on their own for the best price. However, market

[123] This issue is closely related to the effects of arbitrage. As discussed in detail in Section E.2(a), *infra*, arbitrage cannot solve the problem of fragmentation because *the arbitrageur must incur the costs of illiquidity* when attempting an arbitrage. Thus, while arbitrage indeed aligns prices across markets over longer time horizons, it will not solve the

fragmentation may cause trade-throughs for several reasons. First, in a fragmented market, the search cost is higher: the trader has to search for the best price in a number of markets rather than in a single central market. In a fast-moving market, searches across multiple markets may result in a high delay cost. Thus, the trader may decide to give up the search and trade in the market he accessed first.

Second, in cases like the trading of large blocks, it may be in a trader's best interest to trade at a price which is inferior to the best quote due to differences in quote sizes. Suppose, for example, that the size of the best bid in market ***A*** is insufficient to absorb the entire block. The trader may then go to market ***B***, where a larger-size bid is available at a price *below* that of ***A***. The trader may still be better off executing the order in market ***B*** due to its size advantage.

Further, there may be a difference between the rules of markets ***A*** and ***B***. For example, market ***A*** may require prompt disclosure of the trade, whereas ***B*** allows delayed reporting. Then, a trader who may want to conceal his trade will prefer to sell the block in ***B*** for a lower price than the best bid in market ***A***. In these cases, a trade-through occurs, violating the law of one price and hurting market integrity.

Finally, trade-throughs can occur as a result of trading rules that do not enable traders to trade against the best bid and offer available. This can happen, for example, on crossing networks, if their pricing rules do not adhere to the law of one price.

Trade-throughs not only hurt market integrity, but also reduce market liquidity and informational efficiency. First, they reduce the incentive to provide liquidity in the form of tight quotes, because traders realize that even if they have the best quote, execution may take place elsewhere. Second, the information content of quotes and transaction price is reduced, making the

---

problem over the relevant short-term horizon. The empirical evidence cited in this Section shows that this is indeed the case.

market less informationally efficient. When transactions are executed at the same time in two markets at two different prices, when the bid in one market is higher than the transaction price in the other, or when the ask price in one market is below the trading price in another market, the markets are informationally inefficient and prices are more uncertain. This hurts the price discovery process, because traders find it difficult to infer what is the "true" market price.[124]

The phenomenon of price disparity has been observed in the trading of German Bund futures on both DTB in Frankfurt and London's LIFFE. For example, Breedon[125] examined the differences between market prices observed at the same time for four futures contracts traded in these two markets. He found price differences averaging about 1.5 basis points (LIFFE's price was higher) that could not be explained by any institutional difference between the markets or the contract features. In addition, there was a great deal of variability over time in the averages of the price differences between contract prices traded on LIFFE and DTB. For example, on some days the average price difference was nearly 3 basis points whereas on other days it was about 0.3 basis points. Naturally, for individual contracts, the variations in price differences were even larger. Such price disparities between contract prices would not be observed had trading been consolidated instead of being fragmented between the two markets.

[124] See Section B.2(b), *supra*, for a discussion of the effect of informational efficiency on market quality.
[125] F. Breedon, "Why Do the LIFFE and DTB Bund Futures Contracts Trade at Different Prices?" Working Paper #57, Bank of England, December 1996.

4. Implications for the Evaluation of Alternative Execution Procedures

The choice among alternative market structures requires a balancing act among the interests and preferences of different market participants. On the one hand, less restrictive AEPs are desired by counterparties who are willing to transact with one another without having to interact with the complete order flow. However, these desires should be balanced against the long-term effects on market quality. In any market, there are situations where two parties may want to engage in a transaction that will benefit them both; but allowing them and others like them to transact away from the main market will hurt the market as a whole. In markets that are not properly designed and regulated, this leads to a reduction in market quality.

The evaluation of AEP's should balance the short-term interests of particular market participants against the effects on the overall quality of the market. These effects were discussed in Section B, which put forth key evaluation criteria (liquidity, informational efficiency, market integrity and product innovation) and in Section C, which reviewed the evidence on the harmful effects of market fragmentation. Since the introduction of an AEP can result in fragmentation, the potential benefits of AEPs to particular market participants have to be balanced against the potential harm to the market. In particular, AEPs have to be designed in a way that minimizes any negative effects on liquidity, informational efficiency, market integrity and product innovation. We discuss each of these briefly below.

(a) *Liquidity*: Market fragmentation reduces liquidity, meaning that overall trading costs increase. These costs include a wider bid-ask spread, greater market impact costs, and higher search and delay costs. The increase in the costs of execution harms *all* market participants and increases the cost of hedging. When the underlying is a U.S. Treasury Security, an increase in the cost of hedging is likely to translate into higher required yields and a higher cost of financing for U.S. Government debt.

(b) *Informational efficiency*: Fragmentation impairs the informational efficiency of the market. Price discovery becomes less effective and price signals are more noisy. Further, fragmentation provides incentives for free-riding and decreases the production of information in the market. This problem is particularly important in futures markets, which serve as key price discovery vehicles for the underlying. Tinkering with the informational efficiency of the futures markets undermines their value as price discovery vehicles.

(c) *Market integrity*: In Section B.2(c), we identified three elements that undermine market integrity[126]: *(i)* trade-throughs, *(ii)* lack of functional integrity, and *(iii)* discrimination among trader classes. Trade-throughs harm market integrity as they violate the law of one price and are inconsistent with best execution. Functional integrity is damaged when orders cannot be executed quickly under certain market conditions: if an AEP negatively affects the feasibility of expedient execution, the market malfunctions and its integrity suffers. Discrimination among different classes of traders includes discrimination in traders' access to execution facilities, the possibility that the orders of different traders will be executed at different prices at the same time, and the possibility that some traders will be endowed with special privileges that enable them to take advantage of others. Any of those violations harms the honesty of the market, causes mistrust, reduces the use (and hence liquidity) of the market and causes potential participants to withdraw.

We recommend that the impact of AEPs should be evaluated along these criteria, and that trading rules and market structures should be designed to minimize the potential negative effects of AEPs on liquidity, informational efficiency and market integrity.

The evaluation of AEPs should consider two additional effects that were discussed in Sections B and C. The first effect is the "race to the bottom": once transaction execution standards have been relaxed for one exchange, other exchanges may be forced to relax their standards even if they did not want to do so in the first place. Given the fierce competition between the futures exchanges, it is difficult for an exchange to adhere to a strict standard when its competitors relax it. If two counterparties have pre-arranged a trade among themselves and they find a legitimate institutional arrangement that allows them to consummate the trade, they have a very strong incentive to implement it. The result is that the exchange with the stricter rules will lose order flow and become less competitive and less liquid. Hence, it is forced to follow suit — or face the consequences of losing the order flow.

The second point to consider is the effect of market structure on the incentive to develop new futures contracts. At present, exchanges that develop successful new contracts have a significant likelihood of maintaining the liquidity they nurtured in those products. This may be a fragile equilibrium, but it sustains a reasonable incentive to innovate even though competing exchanges can imitate the product. If competing AEPs are allowed to siphon off a large fraction of the order flow developed by the innovating exchange (e.g., all block trades), the incentive to innovate will be greatly diminished, and so will innovation.

In what follows, we consider three AEP case studies: Block trading on the NYSE, the interdealer market in U.S. Treasury Securities, and the CFFE.

---

[126] Clearly, these elements are related to one another and to the other criteria.

D. ALTERNATIVE EXECUTION PROCEDURES AND MARKET FRAGMENTATION

In this Section we consider the effects of implementing AEPs that do not adhere to the concept of open and competitive trading in trading markets. The most common AEP occurs in the trading of large blocks, where large market-impact costs may make it too expensive to bring the block to the floor of an exchange. In the equity markets, this is often related to a separation between retail and institutional trading venues, but the phenomenon is relevant in other cases as well. Most of the evidence on this phenomenon comes from the equity markets, where orders for large-blocks of stock can be executed "upstairs." In what follows, we first discuss the tradeoffs involved in block trading, and then present case studies from two trading markets that represent polar approaches to the problem.

1. The Price Impact of Large Block Trades: General Pattern

The execution of large orders, typically coming from institutions and market professionals, gives rise to special problems in trading markets. The basic problem arises from the fact that large block trades can have a significant market impact: the purchase of a large block tends to move the market price up, and the sale of a large block tends to move the market price down.

Figure D-1 shows schematically the effects of a large block sale in a trading market. First, as discussed in Section B.2(a), *supra*, the overall effect of the block sale is likely to be a decline in the price of the instrument. However, in some cases the sale will result in a permanent price decline, in other cases the price decline will be reversed following the sale, and in many cases the sale will have an intermediate effect: the sale will initially depress the price of the instrument but *some* of the price decline will be subsequently reversed.

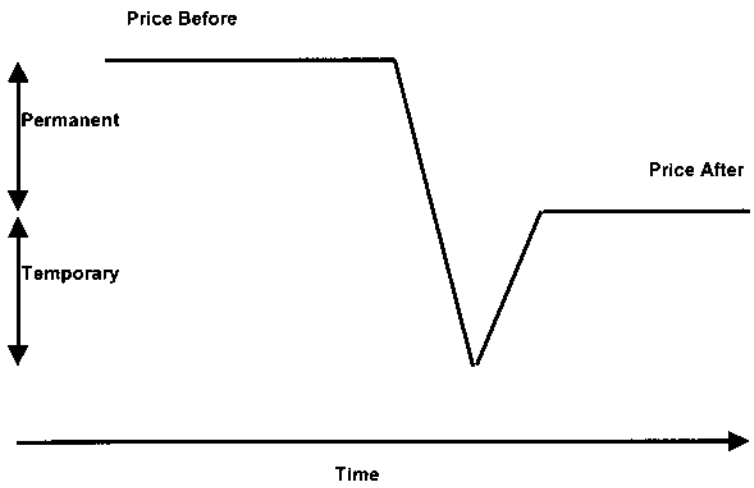


Figure D-1: Effects of a typical block sale.

Figure D-1 shows that on average, the price tends to decline upon the block sale and then it partly rebounds. Thus, the initial price decline due to the block sale is the sum of a permanent component, which represents the transition to the new level of the instrument's price, and a temporary component, which is reversed following the block sale. The permanent component reflects new information, whereas the transitory component reflects an illiquidity (market impact) cost. The magnitude of each component and the adjustment times vary across instruments, but the overall pattern shown in Figure D-1 is quite general. Market impact is measured as a percentage of price; for example, if the price before the block trade was 100, it declined to 99 upon the block sale and then bounced back to 99.75, the total market impact cost was -1%, the temporary component of market impact was -0.75%, and the permanent component was -0.25%.

The price pattern corresponding to a block purchase is similar, except that the price tends to increase rather than decline. Similar to the pattern shown in figure D-1, part of the increase (the temporary component) is reversed and the price settles (on average) at a new, higher level. It will be useful to keep in mind these price effects in the discussion of block trading that follows.

2. *Tradeoffs*

Because the market impact cost discussed above typically increases with trade size, traders have an incentive to break their orders into smaller pieces. However, if all pieces arrive in sequence through the same broker, or if there are other ways for market participants to infer that the pieces came from the same order, the market will still impute an appropriate market impact. For example, Chan and Lakonishok[127] analyzed the market impact costs for institutional investors who execute large packages of stocks over time, and found that the resulting *cumulative* market impact cost of a stock package averaged almost 1 percent for buy packages and –0.3% for sell packages.[128]

In most securities markets, traders are able to arrange large block trades "upstairs." That is, the large order is placed with a broker (or block positioner) who seeks counterparties to the transaction on behalf of the original block trader. The broker may act as principal, buying the block (or part of it) for his own account, or as agent, seeking to place the entire block with other counterparties. In over-the-counter markets, the transaction can often be completed without further requirements, with the possible exception of transaction reporting requirements. In exchange markets, once the block transaction is arranged, it is typically brought to the floor of an exchange for final execution. When a large order is arranged upstairs, the buyer or seller hopes to reduce the

[127] L.K.C. Chan and J. Lakonishok, "The Behavior of Stock Prices Around Institutional Trades," *Journal of Finance* 50, 1995, 1147-1174.

market-impact cost, but he will typically incur higher search and delay costs as well as higher commissions. Thus, the trading of an order upstairs reflects a tradeoff between different costs of illiquidity discussed in Section B.2(a), *supra*: an order will be traded upstairs when its expected market-impact cost is high relative to the added commissions, search and delay costs.

Large blocks of liquid assets can also be traded "downstairs" on an exchange floor, typically after being broken into smaller trades. In this case, the auction market is used directly as the trading venue, resulting in greater immediacy, lower search and delay costs, but possibly higher market impact cost. Downstairs trading of large size can be the result of tight rules against prearranged trades, as is the case on the CBOT and other U.S. futures exchanges, or the result of an optimizing choice by traders who prefer to incur lower search and delay costs while paying somewhat higher market-impact costs. The more liquid the asset and the smaller the order size, the lower the market-impact cost. As a result, more trading should take place downstairs for more liquid assets and for smaller orders.

On the NYSE, slightly more than half of the trading volume is in large blocks, defined as transactions of 10,000 shares or more or stock having a market value of $200,000 or more (whichever is less). The percentage of block trades has grown, along with the increased importance of institutional investors, to more than 50% of share volume in the 1990's.[129]

Examining all orders for the nearby Treasury Bond futures traded on the CBOT on July 30, 1998, we found that 58% of the trading volume involved orders for 100 contracts or more, and 37% of the trading volume involved orders for 200 contracts or more. Thus, orders of 100 contracts or more account for a substantial share of trading volume on the CBOT.

---

[128] The +1% for purchases and –0.3% for sales represent the total market impact from the open on the package's first day to the close on the package's last day.

[129] In 1970, block trades accounted for about 15% of NYSE trading volume. The percentage increased to 50% by 1984. In the nineties, the percentage of NYSE share volume accounted for by block trades ranged between a low of 50% in

Madhavan and Cheng[130] studied about 21,000 block trades in Dow Jones Industrial Average stocks on the NYSE, and found that the probability of block trading upstairs is an increasing function of the block size and a decreasing function of trading volume.[131] Madhavan and Cheng also found that for the Dow stocks, about 80% of block dollar volume is traded downstairs.[132] For the NYSE overall, about 73% of block volume is accounted for by "downstairs" trading. On the CBOT, all trades (except EFPs) take place downstairs because CBOT and Commission trading rules do not allow prearranged trades. Large orders (for 100 contracts or more) were typically broken into multiple pieces.

Table D-1 presents statistics on buy orders for the nearby (September) T-Bond futures contract traded on the CBOT on July 30, 1998. Order sizes were categorized according to the number of contracts in the order; we then calculated the average number of trades and the average trade size in each category. As shown in Table D-1, orders of 100-1,000 contracts were broken into approximately 3-5 trades on average; trades for more than 1,000 contracts were broken into more than 7 trades on average. As the order size increases, both the average number of trades and the average size of a trade increase.

1990 and a high of 57% in 1995. That percentage was 51% in 1997. See, New York Stock Exchange, *1997 Fact Book*, at 16.

[130] A. Madhavan and M. Cheng, "In Search of Liquidity: Block Trades in the Upstairs and Downstairs Markets," *Review of Financial Studies* 10, 1997, 175-204.

[131] Measured by the average block volume in the stock.

[132] Madhavan and Cheng, *Op. Cit.* at 185.

Table D-1: Basic statistics for buy orders for T-Bond futures contracts on the CBOT; July 30, 1998.

Order Size (Contracts)	Number of Orders	Average Trades per Order	Average Trade Size (Contracts)	% of Orders	% of Volume
1	1,658	1.00	1.00	11.75%	0.34%
2-10	5,823	1.53	4.30	41.27%	6.54%
11-20	1,851	2.19	10.23	13.12%	6.14%
21-50	2,645	2.21	23.84	18.74%	19.26%
51-100	1,204	2.76	48.80	8.53%	20.10%
101-200	530	3.28	76.55	3.76%	16.63%
201-500	353	3.70	149.11	2.50%	23.25%
501-1,000	43	5.19	243.77	0.30%	6.65%
>1,000	4	7.25	202.71	0.03%	1.10%
Total	14,111			100.00%	100.00%

Figure D-2: Order Size Distribution - Bond Futures, 7/30/98

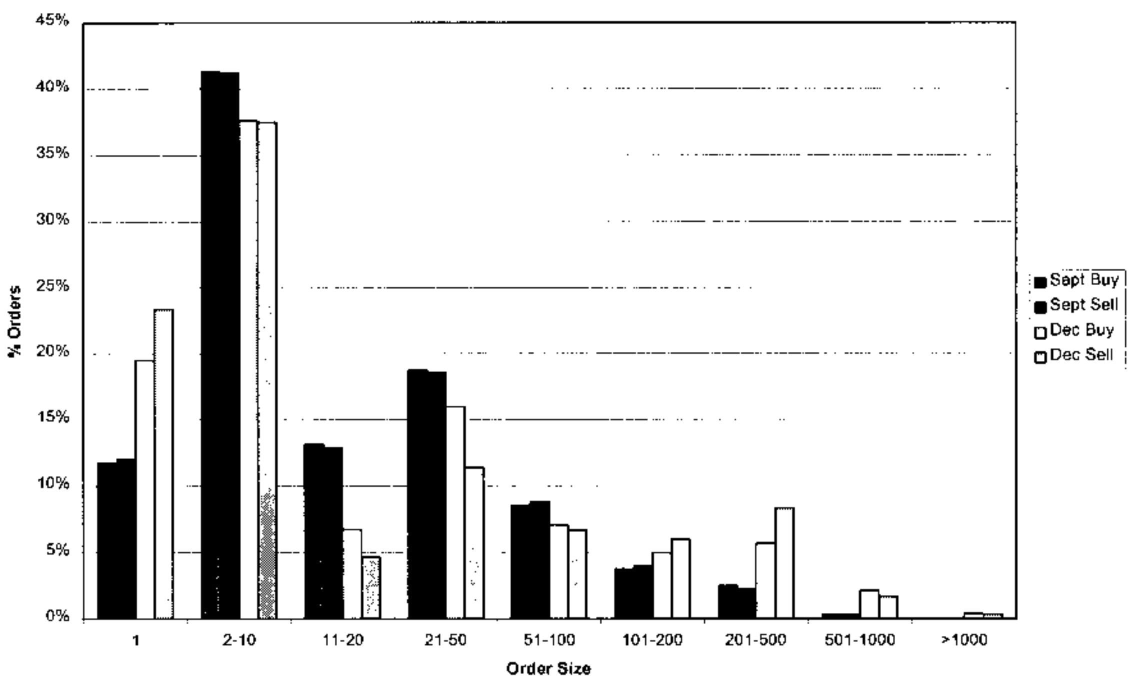
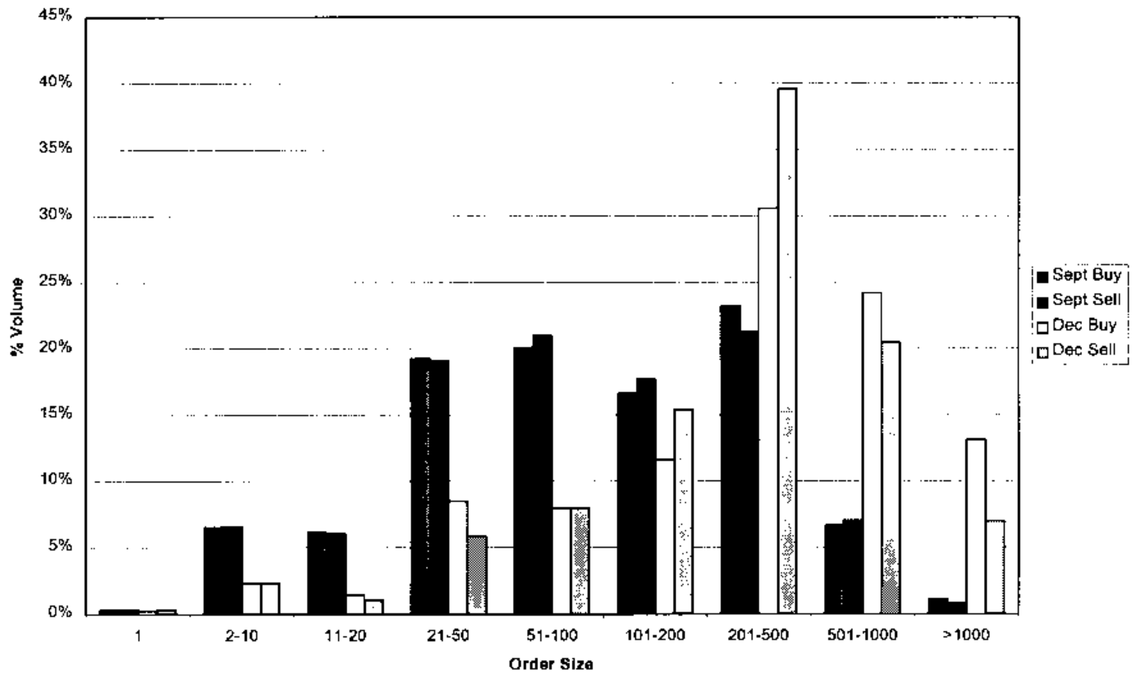


Figure D-3: Volume Distribution by Order Size - Bond Futures, 7/30/98



Figures D-2 and D-3 show how orders and contract volume are distributed by order size. The data consist of buy and sell orders in the September 1998 and December 1998 T-Bond futures contracts that were traded on the CBOT on July 30, 1998. The orders were grouped based on the same order size categories that were shown in Table D-1. For the nearby (September) contract, more than 10% of the orders were for a single contract, about 40% were for 2-10 contracts, and about a third were for 11-50 contracts. By contract volume, Figure D-2 shows that each of the ranges of 21-50, 51-100, 101-200 and 201-500 contracts per order account for about a fifth of the exchange volume.[133] Altogether, orders for 21-500 contracts account for about 80% of the trading volume on the CBOT. Thus, the execution of orders in these ranges is an important issue for the CBOT. For the December T-Bond contract, even more of the trading volume originates as large orders. For example, almost 40% of the December contracts sold were in the 201-500 contract range, and about 20% of the contracts were in the 501-1000 order range.

Keim and Madhavan[134] examined the market impact of block transactions that are upstairs trades. The *temporary* market impact on the day of the transaction due to the block sales was 1.86% for NYSE and AMEX transactions, and 3.28% for NASDAQ National Market System stocks.[135] For buyer-initiated block transactions, the temporary price impact was small. In addition to the temporary price impact, block transactions convey information. Consistent with Figure D-1, there was a permanent price decline following block sales and a permanent increase following block buys, which persisted following the day of the transaction.

Madhavan and Cheng[136] found that 49% of orders for more than 50,000 shares on the Dow stocks are executed upstairs.[137] Of the large (more than 50,000 shares) sell orders—which are more difficult to implement and have a larger market impact[138] — 58% are executed upstairs. For these orders, institutional investors seek the stock market analogue of AEPs, finding that it is in their best interest not to bring the order directly to the exchange floor. While these market participants perceive a benefit from the existence of the upstairs market, an important policy question is how

[133] The percentage is lower (17%) in the 101-200 range and higher (23%) in the 201-500 range.
[134] D. B. Keim and A. Madhavan, "The Upstairs Market for Large-Block Transactions: Analysis and Measurement of Price Effects," *Review of Financial Studies* 9, 1996, 1-36.
[135] See Section D.1 and Figure D-1, *supra*, for a discussion of temporary market impact. Similar qualitative results were obtained earlier by A. Kraus and H. Stoll, "Price Impacts of Block Trading on the New York Stock Exchange," *Journal of Finance* 27, 1972, 569-588.
[136] Madhavan and Cheng, *Op. Cit.*
[137] *Ibid.* at 187.
[138] *Ibid.*

these trades affect the market as a whole. The answer depends on the degree of interaction of the order flows between the upstairs and downstairs markets.

If the order flows in the two markets do not interact (i.e., they are economically separate), our analysis of the effects of fragmentation on market liquidity applies to the two fragmented markets. By our analysis, the likely result is an increase in the overall cost of illiquidity (see Section C.1, *supra*) and a decline in market efficiency (see Section C.2, *supra*) and market integrity (see Section C.3, *supra*): bid-ask spreads are likely to widen, market prices will become more uncertain, and trade-throughs will become common market practice. The liquidity externality (see Section B.3(a)) implies that even if the counterparties to the block trade find its terms mutually beneficial, not exposing it to the rest of the market harms the market as a whole. Furthermore, segmenting the order flow into large and small trades hinders the benefits from mixing trades, one of the key advantages of a central auction market (see Section B.3(b), *supra*).

Although the upstairs market seems to offer higher liquidity for large blocks through search for counterparties, Madhavan and Cheng[139] found that the downstairs market—effectively an auction market—provided competitive prices for most trades, often outperformed the upstairs market, and was a significant source of liquidity. This is because floor brokers in the downstairs market can "work" a large order in the same way as the upstairs brokers, and because of the competition among the crowd on the floor of the exchange.

Madhavan and Cheng suggested that the upstairs market is viable partly because it is being selected by large block traders who can credibly signal that their trades are not motivated by proprietary information. For example, consider an asset manager who manages an index fund that attempts to replicate the performance of the S&P 500 index. When this asset manager sells a block of stock, he is doing so to accommodate the fund's liquidity needs rather than to trade on new information regarding the value of the stock. Thus, the counterparty to the transaction need not fear that the sale is due to an expected decline in the price of the stock, leading to a lower market impact cost. Thus, it turns out that the choice of the AEP-like trading procedure of the upstairs market is not necessarily the result of its superiority as a trading mechanism, but rather a means for uninformed traders to separate themselves from informed ones. But this separation between the informed and uninformed traders, so common between main markets and their satellite markets, is a form of market fragmentation that can lead to an overall decline in the quality of the market (see Sections B.3(b) and C.2, *supra*).

Rules that require the interaction of the order flows in the upstairs and downstairs markets reduce fragmentation and contribute to overall market quality, while placing constraints on how market participants are allowed to execute their orders. By their very nature, such constraints are deemed undesirable in the short term by traders who wish to effect mutually beneficial transactions at terms that they have already agreed upon. But in the long term, such rules are necessary to prevent the harmful effects of fragmentation on market liquidity, efficiency and integrity. The rules should be designed to strike a balance between the desire of the counterparties to the transaction to preserve the terms they negotiated upstairs, and the need to prevent a fragmented marketplace. At a very minimum, trade-throughs should be avoided to preserve the "primary" priority rule of all exchanges: bids and offers at better prices must be executed first. Otherwise, market integrity will be harmed.

Most exchanges go well beyond this minimum requirement and preserve some "secondary" priority rules, such as time priority. Most importantly, the requirement of an "open and competitive market," meaning that the *entire* order flow on one side of the market has to interact with the *entire* order flow on the other side on an equal basis, improves liquidity and price discovery. Currently,

---

[139] Madhavan and Cheng, *Op. Cit.*

most securities markets do not maintain this requirement a strict sense. That is, they make special provisions for block trades, allowing upstairs negotiations and facilitating the crossing of large block transactions. Yet, they also act to preserve the integrity of the auction market while allowing more execution flexibility for the parties to the block transaction.

A key aspect of the problem is that in a competitive environment, exchanges must take into account the effects of their trading rules on the resulting order flow. If an exchange maintains strict trading rules that require large block trades to fully interact with the trading interest on the exchange floor, and an alternative trading venue does not have such requirements, block traders may find it beneficial to divert their order flow away from the exchange floor.[140] For example, U.S. securities broker-dealers sometimes "print" [141] transactions that were negotiated in the U.S. in a less-regulated foreign market to escape U.S. trade-reporting requirements.[142] In other cases, trades are executed by "third market" dealers who are not exchange members, and hence need not adhere to exchange trading rules. Also, trades are sometimes "printed" on a regional exchange to avoid the possibility that part of a crossed block trade would be lost.[143] Since in the short term, large block traders tend to gravitate to the markets that give them the most freedom, the trading rules of trading markets have to be designed so that order flow would not be lost to competing markets. The result can be a "race to the bottom": large-block trades may be lost to markets with looser transaction execution standards unless each market matches these looser standards. In the short term, this reflects rational decisions made by individual traders on a trade-by-trade basis. But in the long term, the quality of the entire market suffers.

---

[140] See discussion in Y. Amihud and H. Mendelson, "A New Approach to the Regulation of Trading Across Securities Markets," *Op. Cit.*

[141] That is, the transaction is reported as having been executed in the foreign market, although it was actually negotiated in the U.S.

[142] *Market 2000 Report*, U.S. Securities and Exchange Commission, 1994, at 59.

[143] Study II, Structure of the U.S. Equity Markets, *Market 2000 Report*, U.S. Securities and Exchange Commission, 1994.

This divergence between the results of rational choice by individual traders and the overall optimum is a direct result of the liquidity externality discussed in Section B.3(a), *supra*. It creates an important role for market regulation—to impose minimum quality standards which all market centers must adhere to. We believe a key role of the Commission is to establish a unified framework for the regulation of AEPs to prevent the "race to the bottom" from taking a toll on the U.S. futures markets. In what follows, we review how a number of markets have balanced these competing interests. In each case, the trading procedure amounts to what the futures markets would consider AEPs.

3. New York Stock Exchange Block Trading Procedures

In our view, the NYSE has managed to accommodate the needs of large-block traders in the face of significant competition. While its large-block trading rules relax some classical auction market standards, large block executions interact in a meaningful way with the exchange's trading floor, preventing market fragmentation. In spite of these rules, the NYSE managed to retain 90% of the large-block trades occurring during regular exchange hours—more than its market share in smaller, retail-size transactions.[144] Block-trading rules with a similar flavor exist in most equity markets as well as on options markets such as the Chicago Board Options Exchange.[145] We review here the rules of the NYSE because of its size and prominence, and because these rules are representative of the rules of other securities markets.

[144] *Ibid.* at 50.
[145] See Chicago Board Options Exchange Rule 6.9.

On the NYSE, all orders are routed to a specialist who controls the limit order book, posts bid and ask prices and manages an active auction on the exchange floor.[146] Historically, much of the liquidity on the floor was provided by the specialist's trades for his own account. As a result, the market impact for large-block trades was substantial. Since the 1970's, the NYSE changed its trading rules to facilitate the pre-arrangement of large block trades upstairs. However, once a large block trade is arranged, it is brought down to the floor of the exchange and executed there following the exchange's block trading rules.

The NYSE does not sacrifice price priority and does not allow trade-throughs: "The highest bid and the lowest offer shall have precedence in all cases."[147] The execution of block transactions is subordinated to modified priority rules that require an interaction between the block trade and the order flow on the exchange floor. NYSE Rule 76 requires that any order in NYSE-listed stock—even a large-block trade that was "crossed" upstairs—must be exposed to the public: "When a member has an order to buy or sell the same security, he shall, except for bonds traded through ABS[148], publicly offer such security at a price which is higher than his bid by the minimum variation permitted in such security before making a transaction with himself. All such bids and offers shall be clearly announced to the trading crowd before the member may proceed with the proposed 'cross' transaction."[149] As a result, the block—even if it was intermediated upstairs, and there are two parties willing to execute it at the proposed "clean-up" price[150]—will be executed first against any outstanding quotes that are better than the proposed price.

[146] Nowadays, much of this auction is managed semi-electronically, as most orders are routed to the specialist electronically through the NYSE's SuperDOT system.
[147] NYSE Rule 71.
[148] ABS is the NYSE's Automated Bond System, which has different rules.
[149] NYSE Rule 76.
[150] The "clean-up" price is the proposed crossing price that was negotiated upstairs.

Prior to executing a large block trade that may not be readily absorbed by the market, NYSE members are required to explore the depth of the market and sometimes consult the specialist.[151] If the proposed "clean-up" price is outside the current quotation, the block trade is a pure agency cross, and the NYSE member determines that there would not be excessive stock loss, then after informing the specialist, "he or she should announce the clean-up price to other members in the Crowd and fill at such price the public orders of other members limited to the clean-up price or better"[152] in addition to filling the reasonable needs of the specialist.[153] Only then is the member allowed to cross the residual block at the clean-up price.

If the member expects excessive stock loss, he may either adjust the clean-up price, or re-announce the clean-up price, inform the Crowd that they would not be given stock at the clean-up price, follow the crossing procedure of NYSE Rule 76, and allow the Crowd and the specialist to provide price improvement over a sufficient period of time. Only then is the member allowed to cross orders for the remaining stock at the clean-up price.[154]

A pure agency cross of 25,000 shares or more may be executed at a price at or within the prevailing quotation if price improvement could not be secured pursuant to NYSE Rule 76.[155]

If the member trades as principal for part or all of one side of the block trade, other NYSE members and the specialist may participate at the clean-up price or better, and the member then proceeds to cross the block following NYSE Rule 76 to provide the opportunity for price improvement.[156] Further, if the member establishes or increases his position in the stock, he must give priority to public limit orders at the clean-up price or better over his own trading in the block.

[151] NYSE Rule 127(a).
[152] NYSE Rule 127(c)(1).
[153] NYSE Rule 127(e)
[154] NYSE Rule 127(c)(2).
[155] NYSE Rule 72.
[156] NYSE Rule 127(d).

In this way, block trades are integrated with the auction market while still receiving special consideration, especially when a large "cross" takes place at or between the prevailing bid and ask quotes. First, the information is brought down to the exchange floor. Second, floor traders and limit orders are offered ample opportunity for price improvement, and trade-throughs are prevented. Finally, the block transaction is promptly reported to the market. The result is that the market for large blocks is not fragmented away from the ordinary auction market. Thus, in spite of the fact that it relaxed its trading standards to accommodate block trades, the NYSE still fares well on the criteria listed in Section C.4.

4. The Interdealer Market in U.S. Treasury Securities

At the other end of the spectrum compared to the NYSE is the interdealer market in U.S. Treasury Securities. The secondary market in U.S. Treasury Securities is an over-the-counter dealer market. Although about 1,700 brokers and dealers trade in the secondary market, about forty primary government securities dealers account for the majority of trading volume. Almost half of the trading volume in U.S. Treasury Securities takes place in the interdealer market, and the vast majority of interdealer trading volume is done through interdealer brokers.[157] Because of the large trading volume and the concentration of the market, the interdealer market in U.S. Treasury

[157] M. J. Fleming, "The Round-the-Clock Market for U.S. Treasury Securities," *Federal Reserve Bank of New York Economic Policy Review*, 1997, July, 9-32.

Securities is, by definition, a professional market that is designed to facilitate the trading of vast amounts of securities.[158]

The bids and offers made in the interdealer market are displayed on computer screens. The rules of operation of the interdealer market were developed for that specific market without regard to trade practice standards for futures or equity securities markets.[159] The "aggressor," who "hits" a bid or "takes" an offer, pays the brokerage commission. Upon execution, the screen flashes to indicate that an execution is taking place. During this time, if say a bid is hit, the bidder gets the right of first refusal on any new offering.[160] This trading procedure, known as "workup," coupled by the anonymity provided by the brokers' screens, is designed to facilitate the execution of large trades: the bidder needs to show only a small amount on the broker's screen, and still gets the right to fill larger amounts at the same price.

This is an effective mechanism for executing large trades among dealers and large market participants. Stigum quoted a broker who explained why automated trading mechanisms that are based on more open and competitive market principles are unlikely to provide a better solution for large trades: "*On workup trades*, the 5s, the 10s, the 25s, and so on, the computer falls short. If you want to buy 1MM and the other guy wants to sell 1MM, the system is fine; but it cannot cope fast with a typical [workup] situation: a guy puts a bid into the screen for 1MM; his bid is hit, and he says [as his successive bids are hit], 'Make it 5,' 'Make it 10,' 'Make it 22.5MM.' The automated system is too cumbersome for that sort of trade."[161] Explains Stigum: "Yet, that is how

[158] Because the U.S. Treasury Securities market is a pure over-the-counter market, it need not have separate AEPs and the criteria of Section C.4 do not apply. However, when these trading standards are inserted into the futures markets, the result is clearly an AEP. See Section E.1, *infra*.
[159] The CFFE's proposed AEPs are designed to mirror the rules of operation in the interdealer cash market.
[160] See, M. Stigum, *The Money Market* 1990, Irwin.
[161] *Ibid.* at 655. Additions are in the original.

large trades typically get done. When a dealer has a sizable job to do, say in coupons, he typically bids for or offers 1MM because he doesn't want to show his hand."[162]

The interdealer market in U.S. Government Securities clearly violates a number of the maxims that apply to organized exchanges and were discussed in Section B. As Stigum puts it, "the government brokers run what amounts to an unlicensed exchange."[163] The trading rules of this "unlicensed exchange" are not subject to the requirements that apply to all national securities exchanges. In effect, this is a pure over-the-counter market that allows trade-throughs, allows unequal treatment of different types of orders and is functionally separate from the retail market (which, by itself, is also a pure over-the-counter market).

The ability to exclude traders from the market managed by an interdealer broker facilitates large block trades but may require traders to go to another market. Explains Stigum: "The practice of giving the last seller first crack at the next bid is one reason why every trader needs to deal with several brokers. Say Merrill has a big selling job to do. They will hit a broker's bid in the issue they want to sell and every bid that comes in after it. Meanwhile, nobody else can hit a bid with that broker for the issue Merrill is selling."[164]

---

[162] *Ibid.*
[163] *Ibid.* at 655.
[164] *Ibid.* at 649.

E. EVALUATION OF THE PROPOSED CFFE

The analysis in the preceding sections provides a framework for evaluating Cantor Fitzerald's CFFE proposal as a case study. The CFFE is designed for the trading of futures contracts on Treasury notes and Bonds of the type that are now traded on the CBOT. Its trading rules implement the "workup" procedure borrowed from the interdealer market in U.S. Treasury Securities[165], and are designed to facilitate the trading of large orders through a noncompetitive trading mechanism. The question, then, is what will be the effects of allowing the operation of this AEP on the overall market for futures on U.S. Treasury Securities? These effects can be separated into *(i)* the interaction between the CFFE and the already-established competitive market of the CBOT; and *(ii)* the performance of CFFE itself.

This Section shows that the AEP proposed by the CFFE will adversely affect the ability of the futures market for U.S. Treasury Securities to serve the national public interests in price discovery and risk management. It will fragment the market and will lead to lower quality execution of traders' orders. This fragmentation will make the futures market less liquid and will hurt the price discovery process, thus lowering the efficiency of the futures market. The lower quality of the market for futures contracts on Treasury Securities will likely spill over into and reduce the quality of the market for cash Treasury Securities as well. This will likely increase the yield required by investors in U.S. Treasury Securities and, as a result, the cost to the taxpayers for financing the national debt.

---

[165] See Section D.4, *supra*.

1. The CFFE is an AEP

The CFFE is fashioned after the system used by Cantor Fitzgerald to trade Treasury Securities in the interdealer cash market: it will use common software, common Terminal Operators and perhaps even the same screens. As discussed in Section D.4, *supra*, the U.S. Treasury Securities market is a pure over-the-counter market whose trading mechanisms do not adhere to established auction market standards. The "workup" procedure which is the centerpiece of that market is clearly designed to facilitate large trades by giving the first best bid or offer the exclusive right to trade additional quantities. Stigum's[166] analysis, cited in Section D.4, *supra*, clearly supports this conclusion.

There should be no doubt that the CFFE is designed to facilitate block trades. In citing the advantages of the CFFE in the conclusion of their paper, Urtestky and Weber cite *first* the ability to trade large orders: "the CFFE should enable large orders to be completed at competitive prices without undue impact."[167] Indeed, the design features of the CFFE favor large orders to the exclusion of other orders even if they are superiorly priced. While, as discussed in Section D, *supra*, block trading does not require such non-competitive practices, this is how block trades are executed in the interdealer market for U.S. Treasury Securities. Stigum's example cited in Section D.4, *supra*, shows how the workup procedure enables traders in the cash market to do large trades, starting with a small amount (1 MM) and then proceeding to successively increase the amounts to 5 MM, 10 MM, 22.5 MM and so on.[168] As Stigum points out, when a dealer has a sizable order, he typically bids or offers only 1 MM because he doesn't want to "show his hand," and he then uses

---

[166] Stigum, *Op. Cit.*

[167] M. Uretsky and B. Weber, "Next Generation Trading in Futures Markets: An Overview of the Cantor Financial Futures Exchange," white paper, June 8, 1998, at 18. They then proceed to state that "In addition, the system offers a number of advantages..."

[168] Stigum, Op. Cit., at 655.

the workup to trade the large block he really wants to trade.[169] Accomplishing this objective requires the exclusion of other traders during the workup — an inherently non-competitive practice that was developed in the pure over-the-counter cash market and violates the most basic principles of trading in an open and competitive market.

The CFFE violates the common standards of an open and competitive marketplace in four ways. First, during the "Exclusive Time," the First Best bid and offer can exclude other traders from trading, *even if they make superior bids and offers*. The "Exclusive Time" is defined as follows:[170]

> "'Exclusive Time' means the limited period of time during the Execution Time, as set forth in the Rules relating to the applicable Contract, when certain traders obtain exclusive rights to trade with each other or with other traders who wish to buy and sell at the price prevailing during the Execution Time."

This means that during the Exclusive Time, the auction market is suspended while transactions are being executed on the CFFE, and other traders are excluded from the trading process. This is a clear violation of the most basic requirement for an open and competitive market.[171]

Second, the CFFE's Market Crossing session executes orders at a randomly-selected price during the three-minute interval following the initiation of the session. This means that even if a small order could be executed on the CFFE at a superior price, the trading rules of the Market Crossing session require its execution at the randomly selected price. This is a clear violation of the requirements of an open and competitive marketplace.

---

[169] *Ibid.*
[170] CFFE Rule 303(a)(2).
[171] See Section B.2(c), *supra*, regarding the effects of trade-throughs on market integrity, liquidity and informational efficiency.

Third, the CFFE does not prohibit transitory EFP's. A transitory EFP is one where the cash leg is simultaneously (or immediately) offset, leaving an outright futures position. "For example, party A purchases the cash commodity from party B and then engages in an EFP whereby A sells the cash commodity back to B and receives a long futures position. As a result of this integrated transaction, the parties acquire futures positions but end up with the same cash market position as they had before the transaction."[172] Thus, transitory EFPs are another mechanism to block-trade outright futures away from the central market.

Such transitory EFPs enable traders on the CFFE to effectively replace the strict trade-execution standards enforced heretofore by the Commission by the lenient trade-execution standards of the over-the-counter cash market. The fact that Cantor Fitzgerald is the largest broker in the cash market for U.S. Treasury Securities and the same phone call to a Terminal Operator can be used to effect transactions in both the cash and futures markets thus makes it easy and convenient to violate even the trading standards of the CFFE.

Fourth, in order to facilitate large trades, the trading rules of the CFFE discriminate against smaller trades. This is both inconsistent with the standards of an open and competitive market and is likely to hurt the liquidity of the market. This issue is discussed in some detail in Section E.6.

2. Market Fragmentation

The CFFE is designed to operate independently and in parallel with the CBOT. The CBOT's pit trading takes place between 7:20 a.m. – 2:00 p.m. Central Standard Time ("CST") (last trading day: 7:20 a.m. to 12:00 noon CST). The CBOT's electronic trading system, Project A, currently operates between 2:30 – 4:30 p.m. CST from Monday to Thursday and between 5:55 p.m. and 6:45

---

[172] Release at 3713.

a.m. CST on Sunday to Thursday, and will soon be extended to trading virtually around the clock. The CFFE is planning to trade from 6:30 a.m. to 4:30 p.m. CST (last trading day: 6:30 a.m. to 12:00 noon CST). This means that trading on the CFFE will largely overlap both the CBOT's pit trading hours and Project A.

The CFFE is not designed to have convenient two-way communications with CBOT, which would enable information to be transferred between the two markets, nor is there a trading protocol that will ensure an adequate degree of integration between the two markets.[173] This will cause market fragmentation and will reduce market quality. As discussed in Section C, *supra*, market fragmentation lowers liquidity, reduces informational efficiency, hampers price discovery and hurts market integrity.

This is not to say that the introduction of *any* new futures market that operates side-by-side with an existing market is necessarily harmful. A properly designed trading system will minimize the negative effects of fragmentation, and will add competition to the marketplace by offering traders a choice between execution systems. However, the new trading system should be designed so as to prevent the fragmentation of trading.

By design, when a trader enters the CFFE trading process, he or she is disconnected from the main trading arena in the CBOT. Similarly, traders on the CBOT are disconnected from the trading process on the CFFE. The order flows in the two markets are separate and do not interact, which is what market fragmentation is all about. There is no mechanism that exposes the orders of traders on the CFFE to the pit traders in the CBOT and enables the latter to trade against CFFE quotes. Nor is there a mechanism that shows the quotes of pit traders to those in the CFFE. The CFFE is designed so that traders with the First Best bid and offer—those who first posted the best quotes on the CFFE—have an *exclusive right* during the Exclusive Time to trade with parties on the other side.[174] During this time, the traders in the system are separated from the rest of the market, thereby excluding any others who are willing to enter a better bid or offer.

While a trader in any market can seek information and possibly trade in the other market, it takes time and effort. In other words, the CFFE increases the *search and delay* costs, which are important components of illiquidity costs. As pointed out in Section B.2, *supra*, traders can avoid search and delay costs by executing orders at the available quoted prices, which may be inferior to the best prices that they would be able to obtain in a consolidated market. As a result, fragmentation forces traders into inferior executions. The search and delay costs are particularly high in a market with high volatility, as the futures market often is. Thus, traders would be more willing to settle for trading at inferior bid and offer quotes that are immediately available in their market, in order to avoid search and delay costs.

Consequently, those who quote prices in one market, knowing that it is costly for traders to compare prices with the other market, will tend to quote worse prices (lower bid or higher offer prices). In addition, they will be willing to absorb large orders only with a greater market impact in any of the two markets. This translates into lower liquidity and wider effective bid-ask spreads resulting from market fragmentation.

One might argue that these problems will be solved by arbitrageurs who have an incentive to align price disparities between the two markets. We address this point next.

---

[173] For an analysis of the effects of arbitrage, see Section E.2(a), *infra*.
[174] *The Algorithm*, CFFE Application, January 7, 1998.

(a) Arbitrage

In all trading markets, a disciplining force that tends to reduce price disparities across related instruments or markets is arbitrage. That is, if the price of a financial instrument in one market, say ***A***, is sufficiently greater than its price in another market, ***B***, there are arbitrageurs who would take advantage of the price disparity. By shorting the instrument in market ***A***, where it is overpriced, and taking a long position in market ***B***, the arbitrageur makes a profit while inducing price convergence between the two markets. Thus, one might argue that it is sufficient to let arbitrageurs prevent the negative effects of fragmentation. This point of view, however, ignores the nature of illiquidity costs that inherently reflect temporary gaps between supply and demand: if arbitrage could close these gaps costlessly and instantaneously, illiquidity costs would not exist in the first place.

It is important to distinguish between longer-term and short-term price disparities. While the arbitrage argument is viable in a long-term analysis, it ignores the fact that illiquidity costs result from short-term gaps between demand and supply, and the associated price disparities are — by their very nature — of the same order of magnitude as the illiquidity costs that an arbitrageur will have to incur in trading across the two markets. Just as arbitrage cannot erase the gaps between demand and supply *within* a given market, it cannot eliminate price gaps *between* different markets. As long as the price gap is narrower than the total cost of two transactions across the two different markets, including bid-ask spreads (as the arbitrageur will have to quickly buy in one market and sell in the other), search and delay costs, exchange fees, execution and position risks[175] and the arbitrageur's own opportunity costs, there is no incentive to engage in such arbitrage.

[175] See section B.2(a), *supra*. Typically, the arbitrageur will have to take a position in one market before the transaction in the other market can be fully consummated.

Another way to recast the analysis is to note that fragmentation results from the fact that only partial order-flow information is available in each market. As Grossman[176] points out, classical economic theory assumes "that all individuals continuously participate in one giant market where they can express their demands for all assets simultaneously as a function of a giant price vector. This assumption of simultaneous and continuous participation in all markets is inconsistent with two important facts: first, it is costly for an individual or an institution to continuously express demands in any single market, and, second, it is simply impossible to trade in all markets simultaneously."[177]

Thus, price gaps between different markets are the necessary result of incomplete information about order flow. For example, across markets like the S&P 500 futures market and the NYSE, "these two markets cannot be perfectly arbitraged at each instant in time because a trader on the floor of one market does not know whether the unusual order flow that he faces is common to both markets or special to his own."[178] Thus, whereas a naïve analysis that assumes the simultaneous and continuous participation of all agents in all markets might suggest that arbitrage will close intermarket price gaps, these gaps will persist whenever order flows are informationally fragmented across different markets.

Because the cost of executing arbitrage transactions is driven by, and is inherently similar to, the cost of illiquidity, arbitrage cannot eliminate illiquidity costs. Market-makers[179] effectively play a similar role to arbitrageurs (buying low, at the bid price, and selling high, at the ask price), but the costs of market-making imply that this form of arbitrage does not extinguish the costs of

[176] S. J. Grossman, "The Informational Role of Upstairs and Downstairs Trading," *Journal of Business* 56, 1992, 509-529.
[177] *Ibid.*, at 509. Grossman also presents a number of examples showing that these assumptions are inconsistent with common practice in different types of trading markets.
[178] *Ibid.*, at 527.
[179] Recall that we are referring to the economic notion of market-maker as liquidity provider. In the trading pit, for example, locals play this role.

illiquidity. Similarly, arbitrage across markets — which is inherently more expensive — cannot void the illiquidity costs resulting from fragmentation. This is why, for example, price differences between the equity market and the stock index futures market can persist for over 20 minutes.[180] And, unlike the case of the futures and cash markets, fragmentation of the market in a given instrument (or virtually identical instruments) in fact increases the cost of arbitrage by increasing illiquidity costs, thus reducing the effectiveness of arbitrage even over longer periods of time.

It should also be pointed out that from a macroeconomic viewpoint, the costs that arbitrageurs incur in narrowing price gaps when the same asset[181] is traded in two different locations come at the expense of all other market participants. Given that splitting the order flow into two separate trading locations does not generate new economic value, the gain of one group of traders (arbitrageurs) must correspond to a loss for other traders in the market. Clearly, the incentive for these traders to participate in the market will be diminished if their cost increases. Notably, these costs are not readily measured from the transaction prices, but they certainly exist, and they reflect both the total gains from arbitrage and the costs of arbitrage.

Indeed, the analysis in Section C cites numerous studies that found significant costs due to fragmentation in spite of the possibility of intermarket arbitrage. These costs can be measured either through price disparities or, equivalently, through the prevalence of inferior order executions. Blume and Goldstein[182] examined the determinants of order flow for stocks that are traded on both the NYSE and on regional markets. They found that the NYSE quotes are usually better than, or equal to, the quotes on the same stocks in the satellite markets. Put differently, the main market—the NYSE—usually defines the best displayed quote which subsequently determines the market

[180] See the studies by Kawaller, Koch and Koch and by Pizzi, Economopoulos and O'Neill, *Op. Cit.*, and the discussion in Section B.1, *supra*. For an analysis of this phenomenon, see Y. Amihud and H. Mendelson, "Index and Index-Futures Returns," *Op. Cit.*
[181] Or substantively the same asset.

price. But importantly, "Non-NYSE markets obtain a substantial proportion of their total trading volume when both sides of their quotes are inferior to the best displayed bid or offer."[183] That is, *trades were executed off the NYSE when the trading exchange posted* ***inferior*** *quotes to those posted on the NYSE.* And this is despite the fact that, in addition, market orders on the NYSE are executed at a substantial *price improvement* over the quoted prices.[184]

Notably, this is the case even though the prices across the different exchanges are disciplined by an electronic system that was designed to integrate them through intermarket arbitrage—the Intermarket Trading System (ITS). The ITS requires that a public market order submitted to any market on the system be executed at the best price that is displayed system-wide, or at a better price. Then, unless the off-NYSE trading exchange is executing the order at a price that matches the best quote on the NYSE (a practice that is in itself problematic, see Amihud and Mendelson[185]), the order must be sent for execution to the NYSE or to any other market that displays the best quote. The empirical evidence still shows that orders are often sent for execution in an inferior market. Thus, it is quite likely that orders will be diverted for execution on the satellite CFFE market on terms that are inferior to those available in the CBOT.

In summary, the CFFE creates, by design, market fragmentation, which is likely to hurt liquidity and increase transaction costs, as discussed in Section C *supra.*

---

[182] M. Blume and M. Goldstein, "Quotes, Order Flow and Price Discovery," *Journal of Finance* 52, 1997, 221-244.
[183] *Ibid.* at 242-243.
[184] See L. Harris and J. Hasbrouck, "Market vs. Limit Orders: The SuperDOT Evidence on Order Submission Strategy," *Journal of Financial and Quantitative Analysis* 31, 1996, 213-231.
[185] Y. Amihud and H. Mendelson, "How (Not) to Integrate the European Capital Markets," *Op. Cit.*

2. Market Efficiency, Price Discovery and Market Integrity

The fragmentation of the Treasury Securities futures market created by the CFFE is likely to result in *trade-throughs* in both markets, because of the possibility that a transaction price in the CFFE will be outside the best quoted bid and offer prices on the CBOT, and vice versa. By the analysis of Section B.2(c), *supra,* this blurs the information content of quotes and trade prices, hampers price discovery and hurts market integrity. This problem is exacerbated by the fact that the CFFE is designed for the trading of large orders. As a result, the bids and offers quoted on the CFFE are likely to be geared towards the trading of large orders through the "workup" process.[186] In contrast, CBOT quotes are made in an open and competitive auction market, where bids and offers are directly transformed into corresponding trades. Because the two trading mechanisms have different objectives and trading procedures, the quotes made in the two markets are likely to be different.

Price accuracy is paramount in providing information to market participants. If a transaction price is outside of the best quoted bid and offer prices, those who observe the information may wonder what is the correct value of the futures contract: is it within the bid and offer prices, or is it the transaction price? The principle of having a single price in the market for the same instrument at the same time (the law of one price) will be violated.

In Section C.2, we cited extensive evidence on the harmful effects of market fragmentation on informational efficiency and price discovery. These problems are compounded by best execution concerns when a market with inferior quotes can attract a considerable order flow. Blume and Goldstein[187] found that for stocks traded both on the NYSE and on other exchanges (regional exchanges and NASDAQ), in most cases the best quote is available on the NYSE. This is

[186] See Section D.4, *supra*.
[187] Blume and Goldstein, *Op. Cit.*

consistent with evidence provided by Hasbrouck.[188] While in general markets revised their quotes and followed the quote set in the market that initiates a quote improvement, the best quote remained for the longest period in the NYSE without being followed by the other markets. Market fragmentation does not provide the incentive for a satellite market to be aggressive in improving its quotes in order to attract order flow. As Blume and Goldstein show, even with inferior quotes, satellite markets attract trading away from the main market that usually initiates the improvement in quote prices. When fragmentation enables a satellite market to attract order flow, that market does not need to contribute to the price discovery process by posting better quotes. The impact on price discovery is detrimental. While the interests of the market with the inferior prices are not materially hurt, the economy as a whole suffers from having less information and a worse price discovery process.

4. Effects of Order Flow Diversion on the CBOT

As discussed in Section D.2 *supra*, orders for large blocks account for a substantial percentage of trading volume on the CBOT. Based on the buy order statistics for the September bond futures contract traded on July 30, 1998 on the CBOT (Figure D-3), about 30% of trading volume is for orders of more than 200 contracts, almost 50% is for orders of more than 100 contracts, and about two thirds of volume are for orders with more than 50 contracts. For the December Bond futures traded on the same day, about two-thirds of volume come from orders for more than 200 contracts, almost 80% of volume are from orders for more than 100 contracts and about seven-eighths of volume are from orders of more than 50 contracts. Given that the CFFE is

[188] Hasbrouck, *Op. Cit.*

specifically designed to trade large orders, this presents a substantial risk to the CBOT and to the liquidity it provides to the entire market.

As discussed in Section D, *supra*, large block traders have a short-term incentive to go to non-competitive markets irrespective of the long-term impact of their trades. Because of the CFFE's non-competitive trading standards, the advantage its workup process gives to large orders and the ability to handle cash and futures market trades through the same Terminal Operator and in the same phone call, the CFFE could be successful in attracting much of the large order volume now traded competitively on the CBOT. By attracting large-order volume away from the CBOT, the CFFE would then reduce the liquidity available on the CBOT.

Providing liquidity entails a significant component of fixed cost, which must be borne regardless of whether the provider is trading independently of the order flow available to trade. If a liquidity provider does not earn a minimum trading profit that covers his fixed costs, he will withdraw from the market. The reason is as follows. When the fixed costs are significant, the average cost of providing liquidity to the market is a decreasing function of the total order volume available and an increasing function of the number of liquidity providers.[189] Thus, a decline in order volume increases the average cost of providing liquidity services and makes them unprofitable unless there is a corresponding decline in the number of providers.

It follows that, in equilibrium, a decline in the order flow reduces the number of locals and other liquidity providers who stand ready to supply liquidity to the market.[190] This is a direct result of the diversion of part of the order flow from the CBOT to the CFFE, and is similar to what happens in other competitive markets: in the presence of fixed costs, a decline in volume leads to

[189] The reason is that when the order volume *per provider* gets lower, the fixed cost is distributed over a lower volume, hence the average cost increases. The average cost per provider gets lower as the total order flow declines or as the number of providers among whom the order flow is split increases.
[190] For detailed analyses, see, e.g., Grossman and Miller, *Op. Cit.*, Hendershott and Mendelson, *Op. Cit.*

exit of suppliers from the market to the point where price can be equated again to average cost. As discussed in Sections C and E.2, *supra*, the result is clear: not only will the liquidity on the CBOT decline — the liquidity of the entire market will decline as well. And, given that the CBOT will likely serve as a "market of last resort," the effects of lower liquidity provided on the CBOT will be most pronounced in times of stress. These effects are discussed next.

5. The Potential Failure of the CFFE Under Stress

An important dimension of liquidity which is affected by market fragmentation is execution speed. As noted in Section B.2(a), *supra*, speed is particularly important in the futures markets. Grossman and Miller[191] observed that futures markets are the leading examples of markets where the demand for immediacy is high. Hence, "Futures markets are successful precisely for those commodities *and in those time periods* where price volatility, and hence the risk of delaying trading, is high."[192] For this reason, the design of any futures market should ensure that this market can fulfill its important role of satisfying the demand for immediacy, i.e., minimizing any potential delay.

A major shortcoming of the CFFE is its inability to handle a large trading volume in times of stress, that is, in times of high trading volume that is usually associated with high volatility.[193] The speed of execution of trades on the CFFE may be satisfactory in tranquil times, when prices move very little. But when prices fluctuate and change fast, the delay in execution that is caused by the trading mechanism of the CFFE becomes costly for traders, because it increases the delay cost (see Section B.2(a), *supra*). In times of stress, the speed of execution on the CFFE is

[191] S.J. Grossman and M. H. Miller, *Op. Cit.* at 619.
[192] *Ibid.*, emphasis added.

expected to be significantly inferior to the CBOT's ability to handle a large trading volume and will cause queues in the market. This is because trading on the CFFE is carried out *sequentially*, whereas trading on the CBOT can be done *in parallel*, and thus a large order flow can be executed in a very short time. Grossman and Miller highlighted this characteristic of the futures markets:

> "The sustained demand for hedging and hence for trading futures quickly is often accommodated by designating a specific physical market place or exchange in which many competing market makers can offer their services *simultaneously*."[194]

The market structure of the CBOT indeed satisfies this requirement, and the very design of the CFFE assumes the existence of the CBOT to handle large trading volumes in times of stress. However, the CFFE may diminish this ability of the CBOT, leading to a situation where the futures markets will not be able to cope with a large trading volume.

Just like its counterpart in the cash market,[195] the CFFE's trading process facilitates block trading at the expense of speed for the rest of the market. Traders access the CFFE by a telephone and need to communicate with a Terminal Operator to arrange the trade. As a result, the CFFE is significantly slower than electronic exchanges like Project A or MATIF's NSC-VF, where traders can accesses the system *directly* through a terminal and execute a trade instantaneously by clicking on the screen. And, as discussed below, the traditional trading pit can accommodate very high throughputs for all orders; the development of electronic order routing facilities is further speeding it up.

---

[193] See empirical evidence for six futures contracts on agricultural commodities, in A.G. Malliaris and J.L. Urrutia, "Volume and Price Relationships: Hypotheses and Testing for Agricultural Futures," *Journal of Futures Markets* 18, 1998, 53-72.

[194] Grossman and Miller, *Op. Cit.* at 619 (emphasis added).

[195] See Section D.4, *supra*.

Further, like its cash market counterpart,[196] the CFFE's workup procedure sacrifices the timeliness of small trades in order to facilitate block trades. During the CFFE's *Exclusive Time*, the First Best bid or offer has absolute priority and can exclude from the market all competing bids or offers—even if they have a *better* price. Meanwhile, *other traders on the same side of the market have to wait*. The costs of queuing are particularly high in times of high volatility. On the CFFE, orders may be waiting for execution on both sides of the market, willing to trade at the same price, but they cannot because of the *sequential* nature of trading. This can become extremely costly during periods of stress.

In contrast, in most auction markets and electronic exchanges orders can be traded in parallel or within a very short time. In the trading pit, even a trader with a large order can execute it against a number of counterparties within a very short time. Seeing the traders in the pit signaling their quotes, a trader or his broker can sell an entire block within a second to a number of locals in the pit, splitting it among them and simply pointing at them and hitting their bids. Importantly, while this trader is carrying out the transaction, others can trade with other traders in the pit *at the same time*. It follows that within one second there may be a number of transactions taking place in parallel by a number of traders and brokers against the quotes of other traders in the pit.

Indeed, the evidence shows that in time of high volume, trading on the CBOT is very fast. October 3, 1997 illustrates an active day where the CBOT was able to handle a large trading volume that may not be possible on the CFFE. The peak 1-minute trading rate for the active T-Bond futures contracts on that day was 512 trades per minute, and there were more than fifty

---

[196] See Section D.4, *supra*.

minutes with more than 200 trades per minute.[197] This speed of execution is impossible on the CFFE, where the trading speed is many times slower because of the ability of a trader with Exclusive Time to lock trading.

The introduction of the CFFE will reduce the ability of the CBOT to handle a large order flow in times of stress. In tranquil times, with low volatility and low volume, the CFFE is likely to attract large-order volume away from the CBOT. In times of stress, however, given the CFFE's shortcoming in handling large trading volumes in a short period of time, it is expected that traders will seek to execute their orders in the CBOT and enjoy the greater speed provided by that market. Yet, as discussed in Section E.4, *supra*, liquidity providers will leave the CBOT once order flow has been diverted to the CFFE. The liquidity providers who left the CBOT because of the reduced order flow in tranquil times will not be available in times of stress to provide liquidity and handle an increased order flow. Altogether, while futures trading may be sustained in tranquil times by splitting the trading volume between CBOT and the CFFE, in times of stress the futures market may be unable to cope.

Any trading market adapts its capacity to provide liquidity to a level that is economically sustainable, commensurate with the average order flow routed to that market. When a market loses a substantial fraction of its order flow, its capacity to provide liquidity is diminished, and it cannot serve any longer as a "market of last resort." As a result, the market as a whole is likely to forfeit its ability to cope with times of stress. As discussed in Section B.2(c), this hurts market integrity.

The October 19, 1987 stock market crash provides an illustration of this problem. The NYSE trading system was not designed to cope with a large trading volume. It functioned well while trading was normal, but on the day of abnormal trading volume and high volatility, the

[197] With throughputs ranging from 203 to 400 trades per minute. The peak number of contracts traded in one minute on that day was in excess of 15,000, and there were more than 50 minutes with more than 4,000 contracts traded per

trading system could not cope. The result was delays in execution, inability to access the market and lack of on-line information. This generated panic among investors and exacerbated the situation.[198] Similar problems may arise in a futures market whose trading system is unable to handle abnormal trading and volatility, and by its introduction it reduces the ability of the other market to cope with abnormal trading activity. On a day of need, the same problems observed in the stock markets in the October 1987 crash may occur in the market for Treasury Securities futures if the CFFE becomes a major contract market.

Not only is the CFFE expected to fail to provide timely execution in times of market stress, it is also expected to fail to provide proper price discovery under such conditions. Martens[199] compared the price discovery processes in the LIFFE and in the DTB, both of which trade contracts on the German Bund. Martens hypothesized that the open outcry systems on LIFFE will be more efficient in a fast moving market, and thus will lead the way in the price discovery process. This is because "at the LIFFE traders can change their quotes with a simple hand signal, while for the limit order book at DTB old quotes have to be withdrawn before new ones can be entered. This lack of speed of handling also withholds traders from entering quotes in the limit order book, afraid as they are for the fast moving market" (p. 13). Martens tested this hypothesis using trade data from both exchanges, and the empirical results are supportive: in high volatility periods, LIFFE has the largest share in the price discovery process. DTB contributes most to price discovery only in low volatility periods.[200]

---

minute.

[198] Miller, *Op. Cit.*, suggests that the volume surge beyond the effective capacity of the NYSE was a major cause of the crash.

[199] M. Martens, "Price Discovery in High and Low Volatility Periods: Open Outcry vs. Electronic Trading," Working Paper, Lancaster University, February 1997.

[200] Note that the comparison here applies to price discovery during high and low-volatility periods, not to trading volumes. At the time of Martens' study, the DTB was already a major player in Bund futures.

By way of comparison, the DTB's system provides far more information than the CFFE: Whereas the DTB's screen shows the depth at different prices on the order book, the CFFE shows only the best quote on each side of the market, and the rest of the information is discarded. Further, the DTB is a faster and more efficient trading system than the proposed CFFE. It is completely automated with easy access by traders from anywhere by computer, whereas the CFFE is a telephone-operated system with sequential quotation and execution.

Therefore, not only will the CFFE lag in execution, but it will also lag in the price discovery process. The CFFE is thus expected to *free ride* on CBOT's price discovery process. Given the analysis in Section C.2 *supra* on the negative effects of free riding on price discovery, the CFFE will have a harmful effect on market efficiency. Its expected free riding on CBOT's production of information will reduce the information produced in the futures market. As pointed out in Section B, *supra*, fast and efficient price discovery is one of the key features of the futures market. The CFFE is expected to hurt informational efficiency, especially in times of high volatility.

In summary, in times of market turbulence the CFFE will fail to perform the two roles of any futures market: speedy execution and price discovery. And, it is likely to detract as well from the ability of the CBOT to fulfill these functions in times of stress, when they are most needed and most valuable.

6. The Handling of Small Orders by the CFFE

The CFFE will create a market in futures contracts on U.S. Treasury Securities which, by design, discriminates against small orders. The CFFE is mainly a large-order trading facility that, in effect, would put smaller orders at a disadvantage in the futures market.

To get the full benefits of the continuous trading process on the CFFE, an order must have a size of at least 10 contracts. Smaller orders of less than 10 contracts can trade according to one of the following two options:

(1) Small orders may be submitted to one of the four daily Market Crossing sessions. Under this procedure, orders do not specify prices at which they are to be filled. Rather, the price is determined by *randomly* selecting a price from the transactions that take place on the CFFE during the three minutes following the Market Crossing session.

(2) Small orders may join the First Best bid or offer. By design, these orders will have lower priority than the orders to which they are joined, and they do not qualify for Exclusive Time privileges. Small orders cannot be used to improve the best bid or offer in the same way that large orders can.

Even large traders may be left with an order of a size less than 10 contracts that will create a problem to trade. This applies to both large traders with "odd lots"—orders whose number of contracts in not an exact multiple of 10—where the counterparty trades with them an odd lot, and when the trader has a "round lot"—an order which is a multiple of 10 contracts—that is traded against an odd-lot order. Statistics from T-Bond futures orders illustrate some of these problems. These statistics show that these exclusions apply to a substantial part of the trading in futures contracts.

Table E-1 shows statistics on orders to buy the nearby (September 1998) T-Bond futures on the CBOT on July 30, 1998 and on the corresponding trades.

Table E-1: Buy orders and trades on the CBOT for September 1998 Treasury Bond futures by size on July 30, 1998.

	Proportion of total orders/trades	**Proportion of total volume**
Trades below 10 contracts	59.46%	9.51%
Trades of odd lots (not multiples of 10 contracts)	69.29%	25.55%
Orders below 10 contracts	44.54%	4.43%
Orders of odd lots (not multiples of 10 contracts)	69.35%	38.43%

Table E-1 shows that the majority of the trades, and nearly half of the orders, consisted of fewer than 10 contracts. Importantly, nearly half of all orders were for fewer than 10 contracts, although their share of trading volume is obviously much lower. This suggests that the CFFE provides inferior terms of execution for nearly half of the orders now coming into the established futures market.

(a) The Market Crossing

Execution in the four daily Market Crossing sessions which the CFFE offers traders with small orders (of less than 10 contracts) substantially reduces the liquidity provided to these traders.[201] While batch trading may be attractive to some investors, most investors value the ability to trade continuously during the day. For example, in recent years, the NYSE instituted two post-trading crossing sessions, where investors can match shares at the closing NYSE price,[202] i.e., at a known price. This is superior to what the CFFE Market Crossing offers, because in the NYSE investors know the execution price with certainty. In one session, investors can trade individual shares. In 1997, the average daily trading volume in this crossing session was 358,500 shares per day, which constitutes 0.006% of the daily NYSE trading volume. In the other crossing session, investors can trade baskets of stocks. The average 1997 trading volume in this session was 4.1 million shares, constituting 0.07% of daily NYSE volume.

Although trading at a single price in the NYSE crossing session saves investors the bid-ask spread that they have to pay during the continuous trading, they prefer to trade in the continuous market. Further, investors are willing to pay more for securities that trade in a continuous market. Amihud, Mendelson and Lauterbach[203] examined the consequences on market quality of moving stocks from a periodic batch trading system to a semi-continuous trading system.[204] There was a significant increase in liquidity and a significant improvement in informational efficiency and price discovery upon the move from the periodic batch system to the continuous system. Investors indeed appreciated the change, and were willing to pay about 6% more for the securities that were moved from batch trading to semi-continuous trading. Additional evidence also shows that securities that were moved from continuous trading to periodic batch trading declined in value. The conclusion from this evidence is that investors strongly dislike periodic batch trading.

---

[201] While the Market Crossing is designed primarily for small orders, large orders can also be executed through the Market Crossing.
[202] New York Stock Exchange, *Fact Book*, 1997, p. 29.
[203] Amihud, Mendelson and Lauterbach, *Op. Cit.*
[204] The stocks were traded on the Tel Aviv Stock Exchange (TASE). The CFFE crossing session is inferior to the periodic batch trading system on the TASE, whereas the CBOT continuous trading is superior to what the TASE offered.

The CFFE offers small trades of fewer than 10 contracts the choice between inferior treatment in the continuous trading process and participation in the Market Crossing sessions, which have a number of undesirable features (discussed below). This makes these investors worse off compared to those with large trades, is likely to have a negative effect on their willingness to trade in futures contracts and hurts market integrity (see Section B.2(c)). The Market Crossing sessions designed by the CFFE expose traders to both *price uncertainty* and *execution uncertainty*, and makes it easier to engage in *price manipulation* that may further hurt market integrity.

***(i) Price Uncertainty*:** Investors who submit buy or sell orders to the Market Crossing session do not know at which price their orders will be executed. CFFE Rule 314-B states how the prices will be determined:

> "... at the end of each Market Crossing in accordance with Rule 303-A, the Cantor System will automatically determine a price for each Contract by randomly selecting a trade executed with respect to such Contract during the three minutes immediately following such Market Crossing. The Cantor System will randomly pick one of the six thirty-second intervals during such three-minute interval, and then randomly pick a trade that occurred within such thirty-second interval. The price at which such trade was executed shall apply to all trades executed with respect to such Contract at such Market Crossing. If no trades with respect to a particular contract are executed within the three minutes immediately following such Market Crossing, no crossing price will be established with respect to such Contract."

The problem with this method of price determination is that investors who enter the session have no way of affecting or knowing the price at which their own orders will be executed nor the probability of execution. Investors who enter a batch crossing session at other exchanges, such as the opening session at the NYSE or at the Tokyo Stock Exchange, can choose between entering a market order or a limit order. With a *market* order, their order will most likely be executed, but at a price that is determined by the market. In this case, they get the benefit of certainty in execution while being exposed to price uncertainty—at their choice. If they want, they can enter a *limit* order that specifies their reservation price. They then have the certainty that their order will be executed only at a price better than (or equal to) their reservation price, and they then choose the uncertainty of execution.

In another type of "crossing session," the U.S. Government's auction of Treasury Securities, investors can choose between entering competitive bids, which state their limit price, or non-competitive bids, which assure them of execution but leave the execution price uncertain. However, at the CFFE's Market Crossing, investors who participate with small orders cannot choose; they are all subject to both price uncertainty and execution uncertainty.

Finally, and most importantly, in a fast moving and volatile futures market, three minutes is a very long time, during which prices can move quite dramatically. Consider a small trader who, after having observed the market price, decides to send an order to the Market Crossing. The price is determined in the following three minutes, but the information that was the basis for the investor's trading decision is no longer relevant three minutes later, and the price at which the transaction is ultimately executed may be such that the investor would not have traded at that price to begin with.

To illustrate, suppose that a trader enters a small sell order into the Market Crossing because he anticipates a price decline and indeed, consistent with his expectations, during the three minutes that follow there are many transactions on the CFFE at continuously and sharply decreasing prices. The sale price which is randomly assigned to the sell order may be from the beginning of that three minute interval—that is, a relatively high sale price—or from the end of the three minute interval, meaning the lowest sale price, considerably lower than the price that the seller anticipated. This means that the seller of a small order is not only subject to price uncertainty, but he is seriously disadvantaged.

(ii) Execution Risk. An investor entering the Market Crossing cannot tell whether his order will be eventually executed, and he cannot affect the likelihood of execution. This gives rise to execution risk.[205] As pointed out above, in the opening crossing sessions in the NYSE or the Tokyo Stock Exchange, investors who enter a market order face price uncertainty but almost always have guaranteed execution. In the CFFE Market Crossing, the buy and sell orders are matched and the number of orders that are executed are the minimum of the buy and sell orders. For example, if there are altogether 50 buy contracts and 80 sell contracts in the Market Crossing, 30 sell contracts will remain unexecuted. Not being assured of execution means that the market does not provide its most important role—liquidity. This reduces the quality of the market (see Section 2 above). Those who are subjected by the CFFE trading procedure to this lower quality of market are investors with small orders of fewer than 10 contracts.

In addition, the entire Market Crossing is annulled if it is not followed by at least one transaction in the three minutes that follow, because then the CFFE will be unable to assign a price to the transaction. Here, both buyers and sellers will have their orders unexecuted. This will be the case even if there are traders who are willing to trade. In a well-designed market, investors will be able to execute their orders against the quotes of traders who stand ready to trade. In the CFFE there may be a situation where traders quote bid and ask prices at which they are willing to trade, investors are willing to trade at these price, but no transaction will take place because of the rules that govern trading.

The likelihood of this event occurring is quite high in contracts that are not actively traded, where it is more likely to have a 3-minute interval with no trade. For example, in the CBOT on July 30, 1998, there were 46 3-minute periods during which there was no trading in the December 1998 T-bond futures contract. These numbers pertain to a contract whose trading is concentrated in one market. The fragmentation of the order flow that will result from the introduction of the CFFE will substantially increase the likelihood of the occurrence of a 3-minute interval with no trade. Then, even a contract with a reasonable volume of trading will be subject to the problem of long periods of no trades. If the order flow is split between two markets, the likelihood of a trader to find a counterparty to trade in each market is lessened, and thus the expected time of no trade increases. This is a direct consequence of market fragmentation. The problem of having a high likelihood of no trade will be particularly severe for the CFFE, which is expected to have a smaller order flow than the CBOT.

(b) Joining Small Orders

An alternative method of trading small orders (fewer than 10 contracts) on the CFFE is to add the small orders to the existing orders that define the First Best bid or offer prices, naturally with lower priority. An incoming order that hits the bid or lifts the offer will be executed first against the large order that defines the First Best bid or offer, and the remainder of the incoming order—if any—will be executed against the added small orders. This procedure discriminates against smaller orders, since they cannot get priority in execution by improving the large, First Best bid or offer; they cannot affect the quote and have to wait passively for an incoming large order that will hit them if

[205] See Section B.2(a).

they are buyers or lift the offer if they are sellers; and they cannot be the first to execute against the prevailing bid or offer quote.

For example, suppose that in the CFFE, the First Best bid is at 100.01 and the First Best offer is at 100.02, both defined by large orders of at least 10 contracts. In this case, there is no trade. A buyer with a small order (fewer than 10 contracts) wants to trade *immediately* by paying 100.02, but he cannot. All he can do is add his order to the best bid at 100.01, and he has to wait for this bid to be hit by an order larger than the size of that bid, because the small order has lower priority than the best bid in place. This shortcoming occurs despite the fact that the small order would have been executed immediately—if the CFFE rules allowed it—because it has an effectively higher *price* priority. This situation does not occur if the order is sent to the CBOT, where such an order would be executed promptly. Likewise, in the NYSE, odd lot market orders (of fewer than 100 shares) are executed immediately against the specialist's inventory at the best bid or offer quotes that exist anywhere (as shown in the Intermarket Trading System).[206]

In addition, suppose that the best offer in the market is 100.03, the best bid is 100.01, and the small investor is willing to buy at 100.02. In the CFFE, the small investor cannot improve the best bid of the large investor. The CFFE will post the bid-ask prices as being 100.01\100.03, while if it reflected the traders' preferences, the best quote would be 100.02\100.03. This means that the best bid and offer prices on the market do not reflect the true bid and offer prices that the traders would quote, and the bid-offer spread is *wider* than it should be, if smaller traders were given the same privileges as large ones. Only if a large trader enters with a bid for 100.02 will the small trader be able to join the large order bidding 100.02. But again the small trader is discriminated against the large one. Although a properly designed trading system would give the small trader time

[206] See J. Hasbrouk, G. Sofianos and D. Sosebee, "New York Stock Exchange Systems and Trading Procedures," Working Paper, New York Stock Exchange, 1993, at 44.

priority and his bid at 100.02 would be executed first, on the CFFE the small order is inferior to the large one, even if it was quoted earlier.

As a final illustration, suppose again that the best bid and offer prices are 100.01\100.03 and two small traders are willing to trade at 100.02. Neither can trade and *both* sides are made worse off. All they can do is queue behind the large orders or wait for the Market Crossing session. Both small traders suffer here, because the CFFE is basically a block trading market that is not designed to cater to small traders.

Because of these inherent problems, the design of the CFFE depends upon the trading system of CBOT, which offers small orders price certainty and execution certainty, and offers an open auction with hundreds of competing traders which makes price manipulation difficult. This may suggest that the CFFE's design leads to a type of *cream skimming*: The CFFE will provide its normal order processing services to the easy-to-handle large orders that are fewer in number and large in volume, whereas small orders that are more costly to handle will be effectively pushed to the CBOT, where large and small orders receive an equal treatment in the continuous trading process. The CFFE is therefore designed to profit from large orders and will impose the higher cost of handling small orders on the CBOT.

(c) Summary

By its design, borrowed in large part from the cash market, the CFFE favors large orders. This means that the system is designed to discriminate against smaller orders, offering them inferior terms of order execution. Smaller trades can be promptly executed on the CBOT at known prices with assured execution. The CFFE will thus free-ride on the CBOT, which offers equal treatment to both small and large orders. If the CFFE were the only trading system in the futures market, it would substantially hurt the interests of market participants who require the smaller trades.

Operating side-by-side with the CBOT, the CFFE may effectively skim the larger and more profitable trades. This will translate into higher execution costs for small orders on the CBOT, which will again disadvantage small investors. As discussed in Sections B.2(c) and B.3(b), this will have a significant negative impact on the liquidity and integrity of the market for U.S. Treasury Securities futures. Similar consequences will apply to any other futures market whose contracts CFFE chooses to mimic.