ISSN 1045-6333

AN EXPLANATION FOR THE ANOMALOUS PATTERN OF NASDAQ QUOTATIONS

Zion M. Shohet*

Discussion Paper No. 163

7/95

Harvard Law School
Cambridge, MA 02138

The Program in Law and Economics is supported by a grant from the John M. Olin Foundation.

*J.D., Harvard Law School, 1995, and Winner of Olin Prize for the Best Paper in Law and Economics.

AN EXPLANATION FOR THE ANOMALOUS PATTERN OF NASDAQ QUOTATIONS

*Zion M. Shohet**

This paper examines the recently reported unnatural pattern of quotes and spreads in Nasdaq quotations and assesses allegations of antitrust violations by market makers. It finds that the collusion allegations are not supported by the evidence and proposes an alternative model to explain the behavior of market makers. The model explores the two-tiered pricing structure employed by market makers, consisting of quote-setting and payments for order flow. Although under competitive conditions the two tiered pricing structure can yield efficient retail pricing, in practice, it leads to substantial inefficiencies in the operation of Nasdaq, generated in large part by the practice of extending best-execution guarantees to brokers. It is this NASD sanctioned practice, the paper concludes, that gives rise to the wide spreads and the unnatural pattern of quotes, and not collusive market maker behavior.

*J.D., Harvard Law School, 1995, and Winner of Olin Prize for the best paper in Law and Economics.

I. INTRODUCTION

During the past year, reports about the operation of Nasdaq have consistently made headlines. The ball started rolling in May, 1994, when the results of an academic study seemed to support allegations that Nasdaq market makers collude to maintain wide bid-ask spreads, citing a seemingly unnatural pattern of quotes and spreads. Before too long, a flood of class-action antitrust suits were filed against Nasdaq market makers and their firms. The Justice Department too entered into the playing field, launching its own antitrust investigation into the operation of the market. The publicity seemed to rattle market makers and their self regulator, the NASD. Market makers reacted swiftly by significantly reducing their spreads. The NASD, while issuing vehement denials of the allegations, has been frantically exploring ways to reform the market.

This paper examines the reported pattern of quotes that led to much of the controversy and assesses the allegations of antitrust violations. It concludes that the collusion allegations are not supported by the evidence, and proposes an alternative model to explain the behavior of market makers. The model reveals substantial inefficiencies in the operation of Nasdaq, generated by a variety of entrenched practices, including the extension of best execution guarantees and payments for order flow. It is these practices, the paper concludes, that give rise to the wide spreads and the unnatural pattern of quotes, and not collusive market maker behavior. Section II provides a broad description of Nasdaq, including the various rules and practices that tend to impact the operation of the market. Section III describes the evolution of the recent controversy and summarizes the case of the collusion theorists. In Section IV, the behavior of market makers is analyzed in detail, constructing a coherent view of the operation of the market. The model's predictions are compared with the evidence in Section V, in an attempt to explain the reported quote patterns. Section VI examines the merit of the various predicates for liability for the behavior of market makers. Finally, Section VII offers some do's and don'ts in the area of regulatory reform.

II. A DESCRIPTION OF NASDAQ

Nasdaq is an interdealer quotation system, regulated by the National Association of Securities Dealers (NASD), a Self Regulating Organization (SRO).[1] In operation since 1971, Nasdaq is registered as a national securities association under Section 15A of the Securities Exchange Act of 1934.[2] Competing dealers who make a market in Nasdaq stocks post their prices in a national quotation system, which can be accessed by other dealers, brokers and some investors.

Nasdaq experienced explosive growth over the last 10 years and has recently eclipsed the New York Stock Exchange (NYSE) in trading volume.[3] The average daily trading volume in the close to 5,000 Nasdaq listed companies approaches 300 million shares.[4] In 1993, approximately 500 dealers made a market in Nasdaq stocks, creating altogether approximately 60,000 market making positions.[5]

A. Nasdaq Compared with an Exchange

Unlike the centralized exchanges, which are agency auction markets with a single specialist making a market in each stock, Nasdaq is a dealer market, with a number of different market makers competing for transactions.[6] On an exchange, such as the NYSE, a specialists posts bid and ask prices for a particular stock. The specialist executes transactions either at the posted prices or at more favorable prices if a buyers and sellers can be brought together on such more favorable terms. In Nasdaq, on the other hand, a variety of different market makers post their own bid and ask quotes, which appear on a centralized quotation system.[7] Market makers compete over order flow on the basis of these quotations, executing transactions forwarded to them by brokers at their posted

quotes.[8] Market makers also negotiate transactions with large and institutional customers at more favorable prices than the posted prices.

The differences between Nasdaq and an exchange-based market extend to their regulatory structure. Nasdaq market makers can choose almost at will the stocks in which they trade, and therefore move in and out of stocks with great flexibility.[9] Exchange specialists, on the other hand, are assigned an individual stock, in which they must continuously make a market. In addition, market makers face a less rigid system of trading rules and surveillance mechanisms, as compared with exchange specialists, giving them almost complete control over quote setting and trade execution.[10] But the existence of many market makers in each stock results in market fragmentation. Whereas all the trades on an exchange flow through the same centralized trading system, in Nasdaq, order flow is directed toward individual market makers, who are fully independent of other market makers in the same stock. Therefore, no single market maker has its hands on the pulse of the entire market, making adequate and timely trade reporting critical not only to investors but also to market makers in their attempt to post adequate prices for their stocks.

B. NASD Rules and Commonly Accepted Practices

Describing the features of a dealer market doesn't begin to explain the operation of the Nasdaq market; it is the various NASD rules and entrenched market practices that shed light on dealer behavior and the functioning of the market. This section describes some of the more important rules and practices; how they affect market maker behavior is discussed more fully in Section IV(B).

1. ***Best Execution Rule*** - SEC guidelines provides that "broker-dealers are under a duty to seek to ensure that their customers obtain the 'best execution' of their

orders."[11] This has been understood to mean that a broker-dealers must search for the most favorable terms possible, under the circumstances, for a customer's transaction.[12]

NASD rules do not explicitly state the best execution rule, still they require members to act in compliance with "just and equitable principles of trade."[13] This language, as it pertains to retail transaction, has been interpreted to require members to "use reasonable diligence to ascertain the best inter-dealer market for the subject security and buy or sell in such market that the resultant price to the customer is as favorable as possible under prevailing market conditions."[14] Among the factors that are considered in determining "reasonable diligence" are: "the character of the market for the security, the size and type of transaction, the number of primary markets checked, and the location and accessibility to the customer's broker-dealer of primary markets and quotations."[15]

In practice, broker-dealers behave as if they discharge their best execution duty when they secure the best quoted prices for their retail customers, without attempting to do any better.[16] In particular, small retail orders have very little chance of execution inside the best bid-ask spread. While 21 percent of public trades occur inside the best bid-ask, 91 percent of the trades that occur within the best spread are transaction of more than 1,000 shares.[17] In fact, the Small Order Execution System (SOES), while guaranteeing execution of customer orders at the best bid-ask spread, allows for no mechanism for price improvement beyond the best spread..

2. ***Payment for Order Flow*** - Payments for Order Flow are rebates or kickbacks that market makers pay to retail brokers for directing their order flow the market maker's way.[18] Typically these are cash payments of a few cents a share, made pursuant to specific arrangements between brokers and market makers. Retail brokers are particularly likely to engage in this practice, so that the practice has its greatest impact on retail customers.[19]

There are two strains of payment for order flow practices: the traditional practice of market makers paying for order flow for OTC stocks and the more recent phenomenon of market makers and regional specialist paying for order flow for exchange-listed stocks, thereby diverting transaction volume from the organized exchanges. While it is the later practice that has caught the attention of regulators, most discussions and regulations extend to both versions of the practice.

Addressing the hazards of payments for order flow, the SEC has recently promulgated new rules requiring enhanced disclosure of payments for order flow practices on customer order confirmations, on customer annual account statements and on new accounts.[20] Specifically, the new rules require broker-dealers to disclose on customer confirmations whether they receive payments for order flow. In addition, the new rules require "disclosure of the broker-dealer policies for determining where to route customer orders that are the subject of payments for order flow."[21] The SEC has also proposed rules for comment that would require "broker-dealers to disclose in writing, on confirmations, upon opening new accounts and on annual disclosure statements, ranges of payment for order flow received on a per share basis; and include, in new account documentation and on annual disclosure statements, an estimate of the aggregate amount of payment for order flow received on an annual basis."[22] SEC rules do not distinguish between the two types of payments for order flow.

*3. **Best Execution Guarantees*** - The practice of paying for order flow goes hand in hand with the practice of granting best execution guarantees, to form what is sometimes referred to as a preferencing arrangement. Under such an arrangement, a market maker guarantees the execution of customer trades at the best bid-ask spread, regardless of whether his own quotes match the best spread. The guarantees do not extend to every retail transaction, but are rather targeted to specific brokers, who, in return, promise to direct their order flow to the guaranteeing market maker.[23] Market

makers make such guarantees to broker for whose order flow they pay, ensuring that a broker who routes its entire order flow to a particular market maker does not run afoul of the best execution rule, even when the market maker's own quotes do not match the prevailing best quotes. It is hard to determine what portion of retail order flow is executed pursuant to best execution guarantees.[24] It is understood, however, that most firms that pay for order flow also guarantee executions at the best bid-ask price.[25]

C. Automated Trading Systems

Most retail transactions are forwarded to market makers through brokers for execution at the best quoted prices. Larger customer, however, often negotiate their trades directly with market makers for prices that are inside the best posted quotes. Transactions among market makers too are frequently negotiated at prices inside the prevailing best spread. Traditionally, Nasdaq transaction were routed manually, via telephone. However, two fully automated systems currently route a subset of the Nasdaq volume in each of the transaction types described above. The Small Order Execution System (SOES) processes small retail transactions at the prevailing spreads and SelectNet facilitates negotiated transactions through computer automation.

*1 **Small Order Execution System (SOES)*** - SOES was implemented in 1984, and was designed to provide an alternative to the traditional telephone-based interaction between brokers and market makers, in the execution of small retail transactions.[26] The system allows public customers to enter orders of limited size into the Nasdaq system for immediate execution at the best available prices.[27] To maintain its retail orientation, NASD limits the size of SOES transactions to 500 shares, and promulgates a variety of rules to prevent professional traders from trading in the system.[28]

Participation in SOES was initially voluntary, but, after the 1987 crash, the NASD promulgated rules mandating market maker participation in SOES and penalizing unexcused withdrawals.[29] Market makers who participate and post quotes in SOES must honor transactions that are forwarded to them up to a "minimum exposure limit," set at two times the maximum SOES trade-size.[30] Thus, a market maker must execute, at a minimum, trades of 1,000 shares at its posted quotes. The system allows market makers to refresh their quotes after executing each transaction.[31] So, in reality, a market maker need only execute one trade, of a maximum size of 500 shares, before it is allowed to change its quotes.

SOES has another important feature: it allows the "preferencing" of transactions. This gives brokers the option to "preference" orders to a specific SOES market maker, who prospectively agrees to accept such preferenced orders from the forwarding broker.[32] Market makers who agree to accept preferenced orders are obligated to execute such orders at the best prevailing quotes, even if their own quotes do not match the best quotes. This feature allows brokers to forward their entire order flow to a particular market maker, without risking violating the best execution rule. Unpreferenced orders are randomly routed to any market maker whose quotes match the best available prices.

2. ***SelectNet*** - SelectNet facilitates the execution of negotiated trades. There is no minimum volume requirement to enter trades into SelectNet, but the system tends to be used for larger trades. Dealers and brokers enter orders into SelectNet, and can opt to either direct their orders to a particular market maker or to all market makers in the security. Once an order is entered into the system, SelectNet allows the counter parties to negotiate the terms of the transaction through the entry of repeated offers and counter-offers.[33] Orders initially entered on a preferenced basis to a particular market maker can subsequently be broadcast to all market makers, to the extent that an unexecuted portion remains.[34] In practice, SelectNet has evolved into a tool that allows broker-dealers to

broadcast transactions to selected broker-dealers, while excluding other broker-dealers and the public.[35] Thus, quote transmission within the SelectNet system, which is often in between the best bid-ask spread, is concealed from the public, and has no direct effect on the prevailing bid-ask spread.

III. THE RECENT CONTROVERSY

Nasdaq has come under a vigorous attack when the results of a recent academic study became public.[36] The study supported allegations that Nasdaq market makers collude to maintain wide bid-ask spreads on OTC stocks, spawning more than 20 class-action suits against Nasdaq dealers and their firms, including some of Wall Street's most prestigious investment banks.[37] The Justice Department soon launched its own antitrust investigation into the matter, and, by December, 1994, it has sent civil investigative demands to several brokerage firms, requesting price information and other records pertaining to Nasdaq trading, dating back to 1985.[38] The requested information included a breakdown of each firm's payments for order flow.[39]

The SEC launched its own investigation into allegations of anti-competitive practices in the Nasdaq system.[40] The SEC's review goes beyond the price-fixing and collusion allegations to examine other anti-competitive practices, such as late reporting and dealers' refusal to honor their posted quotes.

The NASD's reaction to the allegations has been adamant. Joseph Hardiman, the president and chief executive officer of NASD, denied the range of allegations, claiming that there is no "history or any evidence whatsoever of so-called collusion or fixing prices in our market place. One it doesn't exist, and two, when you get through it, the burden of proof is on the other side."[41] Still, the NASD formed its own committee, led by former US. Senator Warren Rudman, to conduct a review of the embattled stock market.[42] The

committee is set to issue its own report about the effectiveness of the governance of the NASD by early spring.[43]

A. Antitrust Allegations

The various civil suits and the Justice Department investigation focus on antitrust issues. Specifically, the question is whether Nasdaq market makers collude to maintain wide spreads and thereby obtain excessive profits at the expense of investors. At first, the thought of monopolistic collusion among the 500 or so market makers, more than sixty of which make a market in each of the large Nasdaq stocks, seems far fetched. But the collusion theorists have been able to build a fairly convincing case for their claims, relying, in large part, on the seemingly unnatural pattern of Nasdaq quotes and on Nasdaq market makers' uniform response to the allegations of collusion.

The study that sparked the interest in Nasdaq's operation reported a perplexing phenomenon: 70 out of the 100 largest and most actively traded Nasdaq stocks were almost never quoted in odd-eighths.[44] For example, fewer than 2 percent of the bid and ask quotes of Apple and MCI, two of the most actively traded Nasdaq stocks, fell on odd-eighths.[45] On average, fewer than 4 percent of the quotes for the stocks in the study fell on odd-eighths (if quotes were evenly distributed, 12.5 percent of the quotes would be expected to fall on each of the eighths). In sharp contrast, the study found that, consistent with a random distribution, more than 10 percent of the quotes fall on each of the odd eighths for NYSE and AMEX stocks.[46]

The study also examined the distribution of inside spreads, the spread between the best quoted bid and the best quoted ask in an individual stock. A comparison of a sample of Nasdaq stocks with a comparable sample of NYSE and AMEX stocks yielded dramatically different results. Approximately 45 percent of the inside spreads of NYSE

and AMEX stocks were of $.25, 25 percent of the spreads were of $.125 and close to 25 percent were of $.375. Nasdaq spreads, on the other hand, were distributed in multiples of $.25, the most common of which were $.25 and $.5. Only 10 percent of Nasdaq spreads were of $.125, and less than 5 percent of the spreads were of $.375, a pattern consistent with the small portion of odd-eighth quotes.[47]

The study suggested that market makers avoid using odd-eighth quotes as part of a collusion to maintain wide spreads, utilizing round numbers as focal points to coordinate prices.[48] The study concluded that although it finds no "conclusive evidence of tacit collusion among market makers," there is no other plausible explanation for the avoidance of odd-eighth quotes.[49] The paper further asserts that because of this behavior, a reduction in tick size (say to one sixteenth) would not necessarily lead to a reduction in spreads.[50]

The market's reaction to these allegations bolstered the collusion theory. On May 24, 1994, in a closed-door meeting of the major Nasdaq market makers, Richard G. Ketchum, NASD's chief operating officer, told market makers that the spreads on some stocks were too wide and asked for voluntary action to narrow them, suggesting that otherwise the NASD would be forced to crack down.[51] Just three days later, spreads on three of the biggest Nasdaq stocks, Apple Computers, Microsoft and Amgen Inc., which until then regularly fluctuated between $.25 and $.5, suddenly narrowed to $.125, staying at that lower level ever since.[52] In addition, according to a study of the 50 most actively traded Nasdaq stocks, in the last two months of 1994, spreads were significantly narrower than in the six months prior to the surfacing of the collusion allegations.[53] The average spread for the stocks in the study narrowed from $.29 to $.20, a 29 percent reduction.[54]

A second academic study, by the same authors who sparked the controversy, documents "a sudden and dramatic narrowing of the inside spreads" for five of the most actively traded Nasdaq stocks.[55] The study attributed the decline in spreads to the fundamental shift in the use of odd-eighth quotes.[56] It argued that once odd-eighth quotes

were introduced in late May, 1994, the proportion of one-eighth spreads immediately rose from almost 0 to over 50 percent, lowering the average quoted spread in these stocks by almost 50 percent.[57] This, the study concludes, was a result of "the collapse of an implicit pricing agreement among the market markers to avoid odd-eighth quotes," which was precipitated by the allegations of collusion.[58] According to the study, the abandonment of the implicit agreement was complete for four of the five stocks in the study. In the fifth stock, Intel Corp., the implicit agreement remained intact, the study concluded, except for a lone market marker who began to undercut spreads by using odd-eighth quotes, which caused modest decreases in spreads.[59]

Supporters of the collusion theory use other evidence to bolster their allegations. They argue that the "intense" competition between numerous market makers is more illusory than real because only a fraction of market makers in any particular stock are active at any given moment.[60] For example, of the 38 Biogen Inc. market makers, only 13 were posting the inside bid price at 11:57am, on August 30, 1994.[61] Indeed, an examination of two samples of Nasdaq stocks demonstrates that even under a broad definition of an active market maker - one that confers "active" status on a market maker that averages at least two hours a day at the inside spread, either bid or ask - more than 35 percent of market makers in each stock are inactive.[62] That far fewer market markers are really competing in any particular stock makes the possibility of collusion somewhat more plausible.

B. Other Unfair Practices

At the same time, Nasdaq has come under fire for a variety of other unfair practices. Unlike the collusion allegations, some of these practices are in direct violation of NASD rules.

1. ***Late Reporting*** - Under NASD rules, market makers are required to report trades in OTC securities, executed during normal business hours, within 90 seconds of execution. Such trades are otherwise designated as late.[63] Still, market makers regularly fail to report trades on a timely basis.[64] This failure leads to the withholding of information from the market, precisely the outcome sought by the delinquent market makers who are trying to prevent movements in stock prices before they are able to unload their own positions. The losers are investors and other market makers who may be trading with a counter-party that has superior information.

2. ***Failure to Follow the Firm Quote Rule*** - The firm quote rule requires dealers to honor their posted quotes. Under NASD rules, a market maker is required to execute a transaction, for at least a normal unit of trading, at its posted quotations.[65] Instead, market makers routinely refuse to execute trades at the quotes they post, or "back-away" as the practice is commonly referred to. Close to 5,000 "backing-away" complaints have been filed with the NASD in the first 9 months of 1994.[66] The NASD's response has been exceedingly forgiving; public disciplinary action has not been taken on any of the complaints.[67]

3. ***Trading Ahead*** - Trading ahead, commonly practiced at Nasdaq, occurs when a market maker receives a customer limit order, but, without executing the limit order, goes ahead and trades for its own account at a better price. In effect, when trading ahead, market makers delay the execution of a customer sell order until the best bid, among all market makers, equals the customer's limit price.[68] Similarly, a retail limit order to buy a security will not be executed until the inside offer drops to the limit order's price.[69] At the time the limit order is pending, market makers freely trade for their own account at better prices. What's more, a single market maker may hold on to two equally priced limit

orders, a buy and a sell, without executing them. The practice allows market makers to charge the full spread on all retail transactions, which market makers contend is justly earned compensation for the risks they take in providing market liquidity and holding an inventory of securities.[70]

The legality of this widespread and long-standing practice came into question in the mid 1980's. In *E.F. Hutton & Co.*,[71] the SEC found that trading ahead of a customer limit order creates a conflict of interest between a principal and an agent, and therefore upheld an order disciplining a broker for executing orders for his own account at superior prices to a pending customer limit order, without disclosing the self-preferential treatment to the customer. Yet instead of prohibiting the practice, the SEC emphasized the need for disclosure when a market maker trades ahead of a customer, at superior prices. Recently, the NASD has proposed rules that would forbid market makers from holding customer limit orders, while, at the same time, trading ahead of those orders for their own account.[72] But even this set of rules does not prevent market makers from trading ahead of limit orders that are placed with other market makers.

IV. Analysis of Market Maker Behavior

This section attempts to provide a coherent description of market maker behavior. It begins by identifying shortfalls in the collusion theory and proceeds by proposing an alternative model of behavior. The analysis treats market maker behavior in OTC stocks, in which they actually make a market. It does not address issues arising from trading in listed stocks in the OTC market.

A. The Collusion Theory and Its Shortfalls

The market maker collusion theory seems to go something like this. Market makers wanted to maintain wide spreads in order to reap excessive profits. They therefore implicitly agreed to avoid quoting stocks in odd-eighths, which would facilitate their conspiracy to maintain wide spreads. Once their plan was uncovered, the agreement broke down, at least for the majority of stocks. After the break-down, market makers stopped avoiding odd-eighth quotes and allowed spreads to decrease. Once spreads narrowed, market makers did not flock out of the market, because they could still earn adequate profits, as prior spreads were inflated and provided for excessive profits.

While simple and seemingly consistent, the collusion theory does not explain the behavior of market makers. First it fails to provide a convincing account for why market makers agree to avoid odd-eighth quotes; the assertion that it aids in maintaining wide spreads is lacking, as will be explained below. Second, the theory does not seem to explain how market makers coordinated their collusion, given the large number of stocks, market makers, and the distinct sets of dealers making markets in different stocks. The arguments that, at any given moment, only a portion of market makers in a particular stock are actually active does not seem to make the collusion theory substantially more plausible.

It is argued that market makers avoid odd-eighth quotes because the use of round numbers facilitates price coordination.[73] But the use of reference pricing to coordinate collusion is only necessary when prices are not easily comparable among different competitors. For example, when buyers and sellers are geographically dispersed, transportation costs become an important element of pricing. Since the transportation component of the price is not uniform, in order to standardize their prices, competitors use "basing point pricing," a pricing system that designates hub cities and determines prices at other cities in relation to their distance from the hub city. The price of a particular stock, on the other hand, is uniform across all market makers, obscuring the need to avoid odd

eighths. Furthermore, it seems that market makers are able to maintain wide and seemingly uncompetitive spreads in thinly traded issues even though they are quoted in increments that are small in comparison to the inside spread, suggesting that quote multiples have little to do with the success of maintaining wide spreads.

What's more, avoiding odd eighths is presumably easily detectable by regulators. After all, this is what tipped off the current storm. Market makers, like other colluders, should, generally speaking, attempt to avoid detection. Using a mechanism that is so easily uncovered by outsiders, particularly when it is not necessary to enforce price coordination, does not seem to make sense.

Finally, collusion theorists argue that market makers agree to avoid odd-eighth quotes to ensure that the inside spread is at least $.25.[74] Avoiding odd-eighths would certainly accomplish this, because for the inside spread to narrow, once it is already at $.25, would require it to dip all the way to zero. This phenomenon was apparent before the May, 1994 reduction in spreads. During that period, the inside spreads in Apple Computer and Lotus Development Stocks used to skip between $.25 and $.50, resting on either spread width for approximately half the trading day.[75] The spreads very rarely moved to either $.125 or $.375. But if market maker's goal is to maintain, on average, the largest possible spread width, it would be to their advantage to use odd-eighths. Using odd-eighths would allow spreads to fluctuate between $.5 and $.375, instead of dipping all the way to $.25. Under such a regime, the average spread would be $.438 instead of the then prevalent average of $.375.

Furthermore, collusion theorists fail to explain how agreements are coordinated among a large number of market makers and across an even larger number of stocks. Because the spreads on different stocks are of different widths and because different sets of dealers make a market in different stocks, successful collusion would require a complex web of agreements. Such a system is unlikely to exists without some evidence of the conspiracy or some evidence of its enforcement, neither of which has been detected.

Alternatively, it could be argued that rather than engage in explicit agreements, market makers communicate through signaling in their quote patterns. This is a more plausible scenario, yet the enforcement problem remains. Academics have proposed a range of sanctions that the dealer community can impose against uncooperative market makers. These include market makers: asking brokers to divert customer orders away from the offender; directing undesirable trades, those based on information, to such dealers; and withdrawing business from violators in other areas such as underwriting.[76] However, no direct observations of market maker imposed sanctions have been detected, casting a significant shadow on the validity of the collusion theory.[77]

Finally, collusion theorists do not explain how such price collusion is maintainable when there are hardly any barriers to entry. Entry can occur in two forms: existing market makers can start making a market in additional stocks, and outsiders may enter the market-making business. To register in an additional stock, a market maker merely needs to notify the NASD of its plans, one day in advance.[78] The requirements for entering the market making business in Nasdaq are not significantly more cumbersome. The most stringent constraint is the requirement that market makers maintain a minimum net capital of $100,000, or $2,500 for each security in which they make a market, whichever is less. A conspiracy would explain why market makers do not invade each other's territory by making a market in additional stocks, but it does not explain why outsiders do not enter the market making business in droves, when they are hardly impeded by NASD regulations.

B. The Behavior of Market Makers Examined

1. A Fully Competitive Model - In a competitive market, brokers would seek to route a customer order to the market maker who posts the best price in the stock. So, for

example, the broker would direct a customer buy order to the market maker with the lowest ask quote. When the transaction is executed, the customer pays the ask price and a commission charged by the broker. The customer's cost in executing the trade is the commission paid to the broker and the spread realized by the market maker.

In a competitive environment, brokers compete for customer trades on the basis of their ability to execute trades at the best spread, on their commission rates and on the quality of other services they provide. Since all brokers will execute trades at the best spread, and assuming a uniform level of brokerage services, brokers will primarily compete on the basis of their commissions. Under competitive conditions, commission rates should be just high enough to cover broker costs and to provide for a normal profit.

In a competitive environment, market makers compete primarily on spreads. In setting their spreads, they consider their cost structure and the behavior of other market makers. The costs of market making can be broken down into three components: variable costs, inventory costs and fixed costs.[79] Variable costs are costs associated with each trade, including clearing and paperwork costs. Inventory costs differ from variable costs in that they vary between different trades, depending on the market maker's inventory level and the length of the inventory holding period at that time.[80] The costs associated with holding inventory include price risks, arising from unanticipated changes in price; financing costs, associated with the borrowing and lending necessary to finance inventory; and information costs arising from potential exposure to other traders and investors who posses superior information about the stock.[81] The level of inventory costs for each transaction increases with the length of the holding period, which depends on the stock's trading volume, since it is harder to unload positions in thinly traded stocks. Inventory costs also depend on the volatility of the stock because greater volatility increases the price risks assumed by the market maker. Variations in trading volume and stock volatility explain why, even in competitive markets, spreads vary across stocks. Finally, dealers bear a variety of fixed overhead costs, including salary and rents.[82]

Presumably, market makers set quotes, in part, on the basis of their costs. Each market maker posts an ask price which is higher than what he perceives the true underlying stock price to be (true price), in order to cover his variable and inventory costs.[83] The ask price will always exceed the true price by at least these cost, since otherwise the transaction will result in a loss. But competition with other market makers will drive down the differential between the ask price and the true price so that it is not significantly greater that the market maker's costs. While a limited expansion in spreads may occur because different market makers may have different inventory costs, by and large, in a competitive environment, spreads would be just wide enough to cover dealer costs and provide for a normal profit.

a. Payments for Order Flow in a Competitive Environment - The practice of paying for order flow introduces a new competitive dimension. Rather than compete on spreads alone, it allows market makers to also compete on the basis of payments for order flow. When handling a customer transaction, a broker would seek to route customer orders to the market maker with the best combination of spreads and payments for order flow. For example, upon receiving a customer buy order, the broker will execute the trade with the market maker offering the lowest price, where each market maker's price is its quoted ask minus its payment for order flow. Analogously, when handling a customer sell order, the broker will seek the highest price, where each market makers price is its quoted bid plus its payment for order flow.

Payments for order flow will cause an expansion in spreads. These payments amount to an increase in the market makers variable costs, no different than an increase, say, in the per trade clearing costs. In a competitive market, the practice of paying for order flow will cause spreads to widen precisely by the per share payment for order flow, allowing market makers to cover their costs and earn a normal profit. Regulators find it troublesome that payments for order flow are made to brokers rather than directly to

customers, because customers may be paying higher spreads without receiving an offsetting benefit. But, as long as the retail brokerage business is competitive and in the absence of other constraints on competition, paying for order flow should not disadvantage customers. This is because brokers will ultimately distribute these payments to customers in the form of lower commissions. Brokers' commissions will decrease because payments for order flow increase brokers' per share revenue. Facing competition from other brokers, each broker will reduce her commission rate by an equivalent amount to the payment for order flow, so that commissions again will just cover the broker's costs and allow for a normal profit, no differently than under the fully competitive model.

The customer's cost of trading is still the spread charged by the market maker and the commission charged by the broker. Since commissions have decreased in an equivalent amount to the increase in payment for order flow, customers should be no worse off under the practice. The market maker-broker relationship is analogous to that between a manufacturer and a dealer. A manufacturer can reduce its price by either cutting the wholesale price or by handing out dealer rebates. Either way, as long as the dealer market is competitive, consumers pay the same retail prices. Thus, the practice of paying for order flow, without other impediments to competition, is innocuous.[84]

*2. **The Operation of Nasdaq*** - A variety of Nasdaq rules and commonly accepted practices constrain the competitive model depicted above. The best execution rule and the practice of guaranteeing best execution contribute to much of the deviation from the competitive model. The following section describe the impact of these practices on the operation of Nasdaq.

a. The Best Execution Rule - The best execution rule, which dictates that customer transactions must, at a minimum, be executed at the best posted bid or ask price,

restricts brokers' from shopping around for the best combination of spreads and payments for order flow. Instead, brokers are required to always transact with a market maker who posts the inside spread, regardless of whether the spread, in combination with payments for order flow, is not competitive. This constraint, on its own, does not have a significant impact on customers, because market makers would adjust their spreads and payments for order flow so that they are never excluded from broker consideration when their overall price is competitive. To accomplish that, market makers would offer to pay for order flow only when they are also able to match the best bid or ask price. When they are away from the best spread, they will cut their payments for order flow to narrow their spreads to the level of the inside spread.[85]

Thus, the best execution rule, without more, reduces the relevance of payments for order flow.[86] The rule causes payments for order flow to decrease and, at the same time, for spreads to narrow. But it has no impact on the competitiveness of market makers' overall price. Payments for order flow may, however, have a role if the tick size in which stocks are traded is too large, so that market makers cannot effecitively compete on spreads. In such circumstances, market makers will use the excessive spread profits to pay brokers for order flow. Therefore, the ability to pay for order flow may actually improve the price that customers pay.

b. Best Execution Guarantees - Market makers routinely grant guarantees to particular brokers to execute their retail trades at the best bid-ask quotes, regardless of their own quotes. This practice has significant implications on the competitiveness of the market. It alters entirely the manner in which market makers determine their spreads because it allows market makers to compete for order flow even when they don't post the best quotes, without running afoul of the best execution rule.

(i) The Impact on Market Makers' Behavior - Market makers' behavior is changed in two significant ways. First, if a market maker's spread is not as good as the inside spread, he would not have to match the inside spread in order to compete for retail business. Second, when on the inside spread (either bid or ask), a market maker will have little reason to improve it. This is because the improvement will automatically be matched by other market makers, under their guarantees to execute trades at the best quotes, which will largely offset the increased transaction volume that a market maker generally hopes to generate from improving its price. For example, if the inside bid, the best posted quote of the price market makers are willing to pay for a stock, is $22.25, a market maker would have very little reason to improve it, say to $22.375, because the new quote will automatically be matched by other market makers, so that much of the retail transaction volume generated by the price improvement will likely flow to other market makers. Therefore, the first mover will merely pay a higher price for the stock, without generating much additional volume.

Both factors tend to make spreads less competitive.[87] At the extreme, one would expect spreads to widen indefinitely because market makers would never have any reason to either match or improve on the inside spread. In practice, however, there are limits to how much spreads will widen. As spreads widen, higher transaction costs will deter customers from trading, thereby reducing the overall retail volume in the security to the point where it may hurt the profitability of market makers. Therefore, market makers, much like monopolists, will set spreads at the profit maximizing level, the level in which the increase in profits from any additional widening of spreads is offset by lost profits from the reduction in transaction volume. An alternative way to describe this is by viewing each market maker as a monopolist with respect to the transaction volume from brokers with whom it has best execution guarantees. This view mimics the operation of the market because, in reality, no other market maker can compete for this transaction volume. Such a market maker will figure out the profit maximizing spread with respect to

that order flow, and improve on the inside spread only when it widens beyond the profit maximizing level. Since all market makers will go through the same exercise, the inside spread will rest at this monopoly level.

Spreads, however, are likely to be narrower than in a pure monopoly because not all transaction volume that is executed at the best posted quotes is governed by best execution guarantees. This is because some brokers do not enter such arrangements with market makers with respect to their order flow, and some of the non-retail transaction volume, which is generally not governed by such guarantees, is executed at the posted quotes.[88] To the extent that market makers compete over the volume that is not governed by best-execution guarantees, they will reduce spreads below the monopoly level. How much will spreads dip below the monopoly level depends on the percentage of transaction volume that is executed at the best bid-ask prices and is not governed by best execution guarantees.[89] Thus, spreads will fluctuate somewhere below the monopoly level but above the competitive level, depending on the nature of the transaction volume. At all times, however, there will be an excess spread of varying widths tagged on to the competitive price.

Another effect of best execution guarantees is to increase the relevance of payments for order flow, even under the restrictions of the best execution rule. The widening of spreads would allow market makers to pay for order flow, something which they cannot afford to do when they vigorously compete on spreads. In fact, under best execution guarantees, market makers may find that paying for order flow is a more effective way to increase their transaction volume than improving on the inside spread, because once the arrangements for payment for order flow and best execution guarantees are set, the narrowing of spreads has little impact on their transaction volume.

Best execution guarantees have another more structural effect: they facilitate the operation of order flow arrangements by making variations between a market maker's spread and the inside spread irrelevant for the functioning of such arrangements. The

guarantees allow a broker to direct retail trades to a particular market maker without first having to ensure that the market maker's quotes match the inside spread. The guarantees improve the ability of brokers and market makers to enter long-term payment for order flow arrangements by ensuring that a broker would be able to direct its entire order flow to a particular market maker.

The importance of this structural effect becomes clear when one considers how cumbersome payment for order flow arrangements would otherwise have to be. Without the guarantees and constrained by the best execution rule, brokers would have to direct their order flow to market makers whose price matches the inside spread. But, in practice, any individual market maker's quotes match the inside spread for only a small portion of the trading day. In fact, on average, a market maker spends less than 1 percent of the trading day at both the best bid and the best ask.[90] Therefore, in order to ensure that they are paid for all their order flow, brokers would have to enter into arrangements with a large number of market makers, and, short of entering into agreements with all market makers in a particular stock, brokers will not be able to ensure getting paid for all their order flow. This would seem to defeat the purpose of the practice, which is to create feeder arrangements between particular brokers and market makers. Therefore, by allowing brokers to route their entire order flow to just one market maker and by causing spreads to widen so that market makers can afford to pay for order flow, best execution guarantees make for an extremely powerful tool.

(ii) The Effects on Nasdaq - In itself, the phenomenon of competition shifting from the spread arena to the payment for order flow arena should not be harmful. To the extent that there are enough market makers competing for order flow, they would bid up payments for order flow so as to fully offset the widening of spreads. And to the extent that the retail brokerage business is competitive, these payments would ultimately flow to consumers in the form of reduced commissions. In the manufacturer distributor analogy,

it should not be alarming that the wholesale price of a product increases substantially while, at the same time, manufacturer rebates to distributors increase proportionately, since, as long as the retail business is competitive, consumers will pay the same price.

But this view proves too simplistic because of differences in the manner in which market makers compete on spreads and on payments for order flow. Market makers compete for transaction volume in a two tiered regime: they compete on the basis of payments for order flow and on the basis of their spreads. But there is a timing differential between the two regimes. Market makers enter into long term payment for order flow arrangements with brokers, by setting the per share amount that brokers would receive for routing their order flow. They only subsequently set their spreads, during each trading day, treating their payments for order flow as a given.

This two tiered approach causes some allocational inefficiencies in the determination of prices. The spread is the only component of the price that can be tailored to each retail transaction, whereas the payment for order flow component is predetermined. The spread, however, is not competitive, hovering somewhere below the monopoly level, as discussed above. The amount by which it exceeds the perfectly competitive spread, or the excess spread, is not constant, fluctuating with the nature of the transaction volume. Payments for order flow, on the other hand, are set at a constant rate per share, so they cannot precisely offset the excess spread in each individual transaction. Thus, even if, on average, payments for order flow offset precisely the amount of excess spreads and even if all payments for order flow were ultimately channeled to consumers in the form of reduced commissions, the price of individual transactions will either exceed or fall below the competitive price, depending on the level of excess spread at the time of the trade. This allocational inefficiency may have a significant effect on retail customers who tend not to be repeat players.[91]

There is a more significant problem with the predetermined nature of payments for order flow. Market makers have to prospectively determine the payments for order flow

on the basis of predictions of future spread widths and future trading costs. Such estimates involve a great deal of uncertainty. First, market makers are not guaranteed that spreads will remain wide throughout any contractual period; it only takes one maverick market maker to reduce spreads for the entire industry. Moreover, in order to predict their costs, market makers have to make ex-ante assumptions about the liquidity and the riskiness of the stock, which too involve substantial uncertainty, especially when made prospectively. What's more, since market makers cannot predict with any degree of certainty their levels of inventory and the inventory holding periods during future transactions, the determination of their future costs involves even greater uncertainty.

These uncertainties will cause market makers to be conservative when they prospectively compete on the basis of payments for order flow. First, market makers will expect a higher return on investment for the additional risk they take in prospectively determining their costs. Moreover, for any period in which the wide spread pattern does not break down, market makers will have retrospectively underestimated their future spread revenues, and therefore paid less for order flow than required to offset the excess spread.[92] This is because prospectively market makers will always consider the probability of maintaining the excess spread, for the entire duration of a contractual period, to be less than 100 percent. Thus, payments for order flow, on average, will be lower than the excess spread, and the overall price paid by customers will exceed the competitive price. Customers, however, derive no benefit from the additional risk-taking by market makers, making them net loser from this practice.[93]

There is another efficiency loss from this two-tiered pricing structure. When making best execution guarantees, market makers commit to both buy and sell retail stocks at the inside spread. In practice, however, at any given time, market makers would rather be either buyers or sellers of a particular stock, for the purpose of adjusting their inventories to the desired levels. The evidence supports this assertion. Market makers' quotes match both the inside bid and the inside ask less than one percent of the time,

suggesting that rarely do they seek to attract transaction volume on both sides of the spread.[94] Moreover, active market makers shift between the inside bid and the inside ask approximately once a day, suggesting that their order-flow needs change rapidly.[95] By forcing market makers to be both buyers and sellers of stocks, regardless of their inventory needs, best execution guarantees increase their inventory costs, leading market makers to return even less of the excess spread in their payments for order flow.

The retail customer, however, derives no benefit from this commitment to cover both sides of the spread. The existence of both an inside bid and an inside ask suggests that there are some market makers who prefer to buy stocks and some that prefer to sell stocks at any given point. The retail customer, it seems, would benefit from having a sell order routed to a market maker who prefers to buy the stock rather than to have the order executed by a market maker who commits to cover both sides of the transaction but at a higher overall price.[96]

c. *Limit Order Handling* - The discussion so far ignored the handling of limit orders. The current practice in Nasdaq is to ignore retail limit orders until the best quote, either bid or ask, matches the limit price. Even if there are two equally priced outstanding limit orders, a sell and a buy, they are not executed until they match the best posted quotes. It is obvious that this practice disadvantages customers in that it prevents them from being matched with other retail customers to obtain execution in between the inside spread. But the practice disadvantages customers in another more systemic way: it prevents them from exerting any pressure on market makers' quotes through the placing of limit orders.

Under current practice, a market maker who receives a limit order knows that he will be able to execute the transaction in the future, if the prevailing posted price moves to the price of the limit order. The market maker does not risk losing the transaction if an

equally priced limit order is placed on the other side. At the same time, the market maker can continue trading for its own account at better prices. Thus, the existence of limit orders places no pressure on the market maker to adjust its spread downwards.

On the other hand, a rule that obligates market makers to pair together outstanding limit orders for execution in between the inside spread, may in fact exert pressure to narrow the spread. This is because whenever outstanding limit orders are paired together for execution, a market maker experiences a loss of potential spread revenue equal to the inside spread. Since retail customers are more likely to place limit orders when the spreads are wide,[97] in order to obtain execution in between the wide inside spread, market makers will have an incentive to narrow the inside spread so as to reduce the proportion of limit orders. This effect will tend to reduce the inside spread, thereby reducing the excess spread above the competitive level.

A more restrictive rule that prohibits market makers from trading for their own account at a better price, whenever a limit order is outstanding, will exert an even greater pressure on spreads. Under such an environment, whenever a limit order is outstanding, market makers would either have to execute the limit order or refrain from trading at a better price. Market makers will therefore either execute the limit order in between the inside spread against their own account, or adjust their spreads so that they match the limit order. Either way, the effective spread charged to customer will narrow.

d. Summary - The above description of market maker behavior attempts to provide a coherent explanation for the pattern of wide spreads. It rejects the collusion theory, and shows that market makers behave competitively under the constraints of the Nasdaq market structure, to the extent that they derive no excess profit given their costs and the amount of risk that they assume. Even though market makers do not derive excessive returns under the two tiered regime, customers are substantially disadvantaged by it. Therefore, there are substantial inefficiencies in the operation of Nasdaq.

V. THE MODEL AND THE EVIDENCE

A The Size of Spreads

The above discussion predicts that Nasdaq spreads would be wider than spreads in a fully competitive environment, not because market makers collude to maintain wide spreads, but rather because of practices that make the market inefficient. As discussed in section III(A), the evidence suggests that Nasdaq spreads are indeed wider than fully competitive spreads. The wide spread pattern is also supported by the explosive growth of Proprietery Trading Systems (PTS), computer driven systems that provide an alternative trading facility. Interestingly, PTSs account for 13% of the Nasdaq/NMS trading volume in 1993, but for only 1.4% of the NYSE trading volume. A pattern of wide spreads would tend to explain the disproportionately greater exodus from Nasdaq.

The more interesting phenomenon, however, is the dramatic reduction in spreads that took place after the initial allegations of collusion surfaced in May, 1994. The reduction in spreads, in the eyes of collusion theorists, is a manifestation of the break-down of previously existing collusive arrangements, or, alternatively, of the creation of yet another collusion to bring down spreads in order to avoid the ire of investors and regulators. These hypotheses, however, are only as good as the allegations that spreads were kept wide via a collusion. To the extent that the model disproves the existence of a collusion, it should, of course, also disprove allegations that an existing collusion has broken down.

The key question is whether the reduction in spreads is inconsistent with the model proposed in this paper. It appears that it is not! The revelation of the wide spreads and the allegations of collusion exerted a great deal of pressure on market makers to reduce

their spreads. The pressure consisted of both the threat of liability from the pending law suits and the threat of additional regulation by the NASD. This intense pressure affected the incentive structure of individual market makers, because maintaining their excess spreads would risk provoking regulators into fundamentally altering the structure of the market. Responding to these threats by narrowing spreads does not necessarily seem uncharacteristic of a competitive behavior. At the same time, market makers were accused of a variety of other anti-competitive practices, such as late reporting and backing away from trades, some of which they may well be guilty of. Their purgatorial reaction with respect to spreads may be reflective of a general guilty demeanor rather than an admission that their spread setting behavior was anti competitive.

Moreover, the actual response of market makers does not hinge on the existence of coordinated behavior or on a break-down of previously existing coordinated behavior. The market structure is such that even if only a few market makers decided to reduce their spreads, it would be in the interest of all market makers to follow suit. This is because, under the system of best execution guarantees, market makers are required to execute trades at the best inside spread. So, in reality, it takes only one maverick market maker to reduce the inside spread for all market makers, for at least the periods during which this maverick sets the inside spread. Once a handful of market makers decide to narrow their spreads, the inside spread may narrow permanently. And once the inside spread narrows, it is immaterial whether all market makers follow suit and narrow their spreads; either way, the inside spread will remain narrow. However, other market makers may choose to follow suit and narrow their spreads as well, for purposes of appearances or in order to compete for order flow that is not governed by best execution guarantees. Therefore, the simultaneous reduction in spreads is not an inexplicable anomaly once the collusion theory is rejected.

B. The Pattern of Avoiding Odd-Eighth Quotes

It is more difficult to explain the pattern of avoiding odd-eighth quotes. As noted earlier, it is not likely that the avoidance of odd eighth quotes was part of a collusion to maintain wide spreads. Why then would market makers uniformly avoid quoting some stocks in odd-eighths?

1. The Costs Associated with monitoring and Altering Quotes - One possible explanation is that market makers sought to reduce the frequency at which they needed to alter their quotes. According to a recent study, in one sample of Nasdaq stocks, close to half the quote changes of inactive market makers, market makers that average less than two hours a day at the inside spread, were moves to remain outside the inside spread, and more than 30% of their quote changes were moves from the inside to outside the spread.[98] Even for active market makers more than 45% of quote changes are either to remain outside the inside spread or to move from the inside to the outside.[99] This suggests that both active and inactive market makers are occupied, for at least part of the trading day, with trying to avoid the inside spread.

Why don't market makers who are trying to avoid the inside spread simply post very wide spreads? For example, if the best bid is $21 and the best ask is $21.25, a market maker who is trying to avoid the inside spread can post a bid of $15 and an ask of $25. There may be an appearance problem in doing so because it would seem to other traders that the market maker is simply not interested in making a market in the stock. Appearances aside, the NASD set strict rules requiring market makers to post quotes that are "reasonably related to the market."[100] Failure to do so may result in stiff penalties. If a market maker's quotes are not reasonably related to the market and she fails to re-enter her quotations, the NASD "may suspend the market maker's quotations in one or all securities."[101]

Market makers are also prohibited from posting excessively wide spreads," spreads that "exceed the parameters for maximum allowable spreads" approved by the NASD.[102] In most situations, the maximum allowable spread is "125 percent of the average of the three narrowest market maker spreads in each security."[103] This leaves very little room for maneuvering, and leaves market makers no choice but to post spreads that are not much wider than the inside spread.

The rules' effect may be so harsh so as to cause market makers to unintentionally fall on the inside spread or even to improve on the inside spread by simply failing to change their quotes. Suppose, for example, that in a rapidly rising market the inside bid for a particular stock is $21.25 and that the inside ask is $21.75, comprising an inside spread of $.5. Constrained by the need to post tight spreads, a market maker who would prefer to remain outside the inside spread posts a $21 bid and a $22 ask. At this point, responding to increased demand, the inside bid rises to $21.5 and the inside ask increases to $22. If the market maker does not immediately respond by raising its ask above $22, it will inadvertently be caught on the inside spread. Suppose the stock's price continues to rise, so that the inside bid goes up to $21.75 and, at the same time, all other market makers raise their inside ask above $22, say to $22.25. If the market maker still fails to respond by raising its quoted ask, it will have single handedly and unintentionally cut the inside spread from $.5 to $.25. Such movements in stock prices can occur very rapidly, subjecting market makers to the risk of entering into transactions they would rather avoid.

Indeed, being caught inadvertently in the inside spread is a common phenomenon. In the first nine months of 1994, close to five thousand backing away complaints have been filed with the NASD .[104] Market makers presumably back away from trades when they inadvertently post quotes that are more favorable than they intend. This happens when they are not quick enough to adjust their quotes in response to market movements, as demonstrated above. Presumably, market makers could avoid such glitches if constant watch was kept on the movements of each stock by individuals authorized to enter quotes

into the Nasdaq computer system. But such close monitoring could be very costly. The staggering number of backing away complaints suggests that market makers choose not to maintain such a close watch on their stocks, presumably because of the costliness of such close monitoring.

By posing quotes that fluctuate in fourths rather than eighths, market makers reduce the frequency in which they need to change their quotes. The inside spread will experience approximately half the movements when quotes move in fourths rather than eighths. Thus, market makers who would like to remain on the inside spread would need to make half the adjustments in their quotes in order to accomplish this goal. At the same time, market makers who wish to avoid improving on the inside spread would be able to do so by adjusting their spreads less frequently too. Thus, the practice of quoting in fourths may simply be a practical way of reducing the costs associated with monitoring and altering quotes.

2. *The Accuracy of Quotes* - When choosing the increments in which they quote their stocks, market makers make a trade-off between the accuracy of their quotes and the costs they expend on adjusting their quotes. Their decision to alter quotes in fourths rather than eighths means that they are willing to ignore factors that would normally cause spreads to move by only one eighth. Instead, market makers choose to respond to pressure on stock prices only when it is significant enough to merit a one fourth movement. The larger the increment in which market makers quote their stocks, the less accurately will the stock price reflect the available information on the stocks.

According to the model presented in Section IV, market makers maintain spreads that are substantially wider than the fully competitive spreads, because of the monopoly problem arising from the operation of the best execution rule. The true competitive price of a stock is somewhere in between the posted inside spread. The wider the inside spread, the less apparent is that true price, and therefore the less apparent is the relationship

between information and price movements. Because of this wide spread, information that would normally lead to small fluctuations in the price of the stock may not ever get reflected in the stock price. This means that the maintenance of wide spreads already sacrifices the accuracy of stock prices. Therefore, the decision to quote stocks in fourths rather than eighths may not lead to a much greater accuracy loss. Given the lower monitoring costs and the relatively insignificant accuracy loss associated with larger qouting increments, the decision to quote in fourths rather than eighths may be fully rational. Admittedly, there is a chicken and an egg question here. Collusion theorists would argue that the decision to quote in fourths is the cause of wide spreads, and not a side-effect of wide spreads, and that therefore it is the culprit behind the decreased accuracy of quotes.

3. *The Generation of Sufficient Transaction Volume* - The effects of best execution guarantees and payment for order flow arrangements may have lead market makers to quote stocks in fourths for another reason. When a market maker moves to improve the inside spread, the move is automatically matched by other market makers as to retail volume with respect to which they extend best execution guarantees. The initial mover can capture only a small portion of the increased transaction volume generated by its price improvement, that which is not captured by other market makers' best execution guarantees. In essence, best execution guarantees dampen the transaction volume that an individual market maker can hope to capture from improving the inside spread. Since, by and large, market makers improve on the inside spread to generate transaction volume, they would have to make their improvements in larger increments in order to generate the same transaction volume derived in a fully competitive market. Thus, it is plausible that market makers avoid changing the spreads by increments of one eighth because such changes generate too small an impact on their transaction volume.

3. *Summary* - While none of the above explanations can account alone for the pattern of avoiding odd eighths, all the explanations together provide for a reasonable rationale for this behavior. Of course, these alternatives to the collusion based explanation make substantial more sense once it is accepted that spreads can be maintained at a wide level without a conspiracy among market makers.

C. Market Entry

NASD rules impose no significant barriers to becoming a market maker in a Nasdaq stock. Still, there hasn't been an appreciable increase in market makers over the last five years. The number of market makers has declined between 1989 and 1991 from 458 to 425, and has increased to 492 market makers in 1993.[105] But the changes in the number of market makers seem more related to the number of stocks traded in Nasdaq than to anything else. Indeed, the average number of market makers per security has been largely stable, hovering between 9 and 11 per security over the last five years.[106]

Why hasn't the number increased more appreciably, at least through May, 1994, if market makers were able to set wide spreads and to extract abnormal profits. The question is really two fold: why haven't existing market makers entered new stocks and why haven't outsiders entered the market making business? Under the collusion theory there is an easy explanation to the first question: market makers engaged in complex agreements which divided-up the turf; any deviations from these arrangements would risk sanctions or penalties by other market makers. But there isn't as obvious an answer, or for that matter any plausible answer, to the second question.

Under the model presented in this paper, however, the answer to both question is equally as obvious: each market maker is behaving competitively under the constraints of NASD rules, and therefore outsiders cannot expect to profitably outbid existing players.

This applies both to outsiders who are considering entering the market-making business and to existing market makers who are considering making a market in additional stocks.

Alternatively, collusion theorists point to the absence of an exodus from the market making business after the substantial reduction in quotes as evidence that market makers were making abnormally high profits before the reduction.[107] But this is not necessarily the case. Under the model, market maker's price setting, during any payment for order flow contractual period, factored the possibility of a breakdown in the monopoly price structure. The possibility of such a breakdown is precisely why market makers retained some of the excess profits from the wide spreads and did not distribute them back in the form of payments for order flow. Thus, in the short term, it is reasonable to expect that market makers' coffers have reserves designed to wheather the strom and to honor their payments for order flow obligations. In the long term, however, if spreads remain narrow, market makers will be forced to reduce their payments for order flow, so that they can at least break even. Of course, there is no guarantee that spreads will remain low, in which instance market makers will be able to continue to pay for order flow.

V. LEGAL ANALYSIS

The sixty-four thousand dollar question is whether, under the proposed model, market makers are violating the rules. This section analyzes market maker behavior under antitrust laws and under the agency duties imposed on broker-dealers as to their dealings with customers.

A. Antitrust Analysis[108]

The model of market maker behavior, while rejecting the idea that market makers collude to maintain wide spreads, does not slam the door on all antitrust theories. The analysis predicts that, hard as they might try, justice department attorneys will not find a smoking gun or any direct evidence of illegal communications among market makers. But there is a more subtle interaction which in itself may constitute an antitrust violation. The pattern of wide spreads that characterized Nasdaq quotations could not have been maintained if not followed by all, or at least the vast majority of market makers. Likewise, the avoidance of odd-eighth quotes would not have been sustainable unless practiced by all, or at least the vast majority of market makers. To some extent then, the behavior of each market maker evinces a reliance on the behavior of all other market makers. It is as if each market maker, by maintaining wide spreads and avoiding odd-eighth quotes, is signaling a participation in a common scheme.

The law on such conscious parallelism is very murky. The question is whether such conscious parallelism rises to the level of "contract, combination or conspiracy under Sherman Act §1.[109] It is clear that an explicit agreement is not a "prerequisite to an unlawful conspiracy."[110] A Sherman Act violation may be found when there is "a unity of purpose or a common design or understanding, or a meeting of the mind in an unlawful arrangement."[111] On the other hand, mere conscious parallelism does not seem to suffice. As the Supreme Court put it: "this Court has never held that proof of parallel business behavior conclusively establishes agreement, or phrased differently, that such behavior itself constitutes a Sherman Act offense. circumstantial evidence of consciously parallel behavior may have made heavy inroads into the traditional judicial attitude toward conspiracy; but 'conscious parallelism' has not yet read conspiracy out of the Sherman Act Entirely."[112] It is clear therefore that conscious parallelism, without more, does not give rise to liability.

Courts require something in addition to conscious parallelism in order to impose liability or to find that a claim involves a triable issue of fact. These "plus factors" include

acts against self interest, indications of express collusion and poor economic performance.[113] Poor economic performance, which exists in industries with excessive prices, persistent excess profits and patterns of increased prices even as demand is falling, bears on the existence of a conspiracy,[114] but isn't easily applied to market making because of the lack of profitability information. The other two factors are examined below.

1. Acts Against Self Interest - Conscious parallelism, without an independent business reason for each player's conduct, would likely give rise to liability.[115] At the same time, an independent business reason often tips a court's ruling against finding a violation.[116] The harder question, however, relates to acts that are contrary to self-interest unless rivals behave accordingly. For example, how should an oligopolist's refusal to reduce prices below the monopoly level be treated when he predicates this decision on the assumption that rivals would follow the price reduction and therefore make it unprofitable.[117] The answer to whether such interdependent conduct amounts to a violation is ambiguous.[118] But there is language in some lower court cases that would point to liability: [p]arallel conduct is clearly not sufficient to make out a conspiracy without some showing that the actions taken appear to be contradictory to the self-interests of the parties or that the actions of one party are rational only if the other parties to the alleged conspiracy act in a a similar manner."[119]

Still, market makers seem to have a lot going for them under the self interest prong; it is in each of their best interests to maintain wide spreads and to avoid quoting in odd-eighths, as discussed in Section V(B). The problem is that the majority of market makers need to behave in this way in order for the prevailing spreads to remain wide and for the avoidance of odd-eighths to be sustainable. And market makers must have awareness of the parallel behavior by all other market makers. Still, the behavior of market makers is not as problematic as that of the oligopolist, which itself may not be an antitrust violation, because, unlike the oligopolist, a market maker would be well advised

to follow this behavior even without consideration for the reaction of other market makers.

Market makers do not fail to narrow the inside spread because of the concern that other market makers will follow suit, but rather despite the fact that other market makers are not likely to follow the leader. Thus, the behavior of each market maker does make independent business sense, and that it takes all market makers acting similarly to maintain wide spreads does nothing to extinguish this independent rationale.

The practice of avoiding odd eighth quotes is somewhat more troublesome since it only makes sense when followed by all market makers. Here it really matters what our starting point is. Clearly, in a world in which all market makers avoid odd eighth quotes it is in each market makers independent self interest to continue quoting in odd eighths. But how do we get to such a world? Presumably, at some point, market makers consciously ceased quoting stocks in odd-eighths. This behavior is interdependent in that the first movers must have counted on other market makers to follow. This is analogous to an oligopolist's decision to raise prices, counting on other sellers to follow suit.

But not exactly. Unlike an oligopolist's price increase, it may make business sense for a market maker to avoid odd eighth quotes, even when other market makers continue to use such quotes. This pertains both to market makers who try to avoid the inside spread and to those who try to improve on it. Clearly, when a market maker's goal is to avoid the inside spread he could do so by either quoting in fourths or eighths, without having any impact on its business, so it is not against his self interest to quote in fourths. Moreover, market makers who seek to improve on the inside spread may find it in their best interest to quote in fourths, as long as the spread is wide enough, in order to generate adequate transaction volume.

The pattern of Intel Corporation quotes immediately after the allegation of collusion became public demonstrates this point. Between March 31, 1994 and June 24, 1994, the daily fraction of odd eighth quotes entered by market makers in the stock

fluctuated between 6 tenths of a percent to 3.1 percent.[120] During each of the trading days in the same period, of the 61 market makers in the stock, between 3 and 17 posted one or more odd-eighth quotes. During the same period, the average daily duration that the prevailing bid-ask spread was at one-eighth varied between 0 percent and 31 percent.

Thus, in this period, market makers went back and forth in their use of odd-eighth quotes, and continued avoiding odd-eighths even when as a significant number of market makers used them. To some this may be evidence that these market makers were resisting the breakdown of their collusion. But such a recalcitrant behavior in the face of intense allegations of collusion is unlikely. It is more plausible that, as long as the inside spread remained wide for the majority of the trading day (it remained wide for at least 70% of the trading day during that period), it was in each market maker's best interest to continue quoting in fourths. Thus, the avoidance of odd eighth quotes is not against self interest and therefore should not give rise to liability.

2. ***Indications of Express collusion*** - Evidence suggesting an express collusion between market players is the second "plus factor," which, together with conscious parallelism, often gives rise to liability.[121] This factor is not much different from the independent business rationale prong, since, whenever the rationale applies, the evidence is less likely to suggest an express collusion. Courts find the test useful when the pattern of prices is so improbable that the only plausible explanation is an express collusion.[122] A court may well find the striking pattern of avoiding odd-eighth quotes and the abrupt abandonment of the practice baffling and improbable enough to imply the existence of an express collusion.

This will be unfortunate. To do so is to ignore the truly improbable nature of an express collusion between such a large number of market makers who make a market in thousands of stocks. But, more importantly, it is to ignore the true incentive structure faced by market makers and the real problems that plague Nasdaq.

B. Broker-Dealers' Duties to Customers

Generally, a broker-dealer who acts as a principal in the sale of a security to a customer is limited in the amount of a mark-up that he can charge the customer.[123] Mark-ups above 10 percent were generally considered excessive and therefore triggered liability.[124] The NASD adopted a more restrictive "5 percent policy," which considers mark-ups of 5 percent or above to be unreasonable.[125] The amount of mark-up is computed by comparison to the contemporaneous prevailing market price.[126] When there is an active, independent and reliable market for a security, the representative inter-dealer quotes establish the prevailing price on the basis of which mark ups are computed.[127] Much of the litigation in this area focuses on the determination of the prevailing market price for more obscure securities, in which the best bid-ask prices are not meaningful.

Thus, current doctrine with respect to actively traded securities addresses mark-ups above and beyond the prevailing spreads, rather than tackles anomalies in the spreads themselves. To the extent that market makers do not charge excessive mark-ups beyond the prevailing bid-ask spreads, they would seem to escape liability under current law. But an argument can be made that even for actively traded securities the best bid-ask spread does not establish the representative price on the basis of which mark ups should be computed. The evidence that spreads were excessively wide and that a large portion of interdealer and institutional trades occurred in between the inside spread would seem to indicate that the representative price is somewhere in between the best bid-ask quotes. This may have some support in the case law which requires that the reliability of quotations "be tested by comparing them with actual inter-dealer transactions."[128] Indeed, in stocks that are not actively traded, whose quotes are therefore commonly agreed not to

establish the representative price, the price is often determined on the basis of the market makers contemporaneous costs in transactions with other dealers.[129]

If mark-ups are indeed calculated on the basis of the price of interdealer transactions that is somewhere between the best-bid ask spread, it is likely that some principal trades in actively traded securities before May, 1994 violated the 5 percent rule, particularly when the excessive spread is added to the mark-up charged by the market maker. This aggressive interpretation of the law will allow the SEC to target individual market makers in specific principal transactions for violation of the mark-up policy.

C. The Appropriateness of Using Current Law to Sanction Market Makers

It seems therefore that a very aggressive enforcement of either the antitrust laws or mark-up rules may subject market makers to liability for their spread setting behavior. It is not clear, however, that such an approach is desirable. It would subject market makers to liability for behaving in a way that was not thought to be illegal by regulators and likely by the market makers themselves.

More importantly, holding market makers liable for past behavior will not necessarily delineate the parameters of appropriate market maker behavior for future purposes. Finding that market makers engaged in a tacit collusion tells them little about what it is precisely that they cannot do in the future. How can market makers eliminate the appearance of such vague concepts as conscious parallelism or interdependent behavior? Antitrust liability will therefore be interpreted as a general command to keep spreads narrow, without explaining what precisely it is about the practice of keeping spreads wide that makes it illegal.

Liability for excessive mark-ups will do no better. Although the legal theory is more explicit, it does not target the source of the evil. Similar to antitrust liability,

excessive mark-up liability will restrain market makers from posting wide spreads. Unlike antitrust liability, it provides more explicit parameters for how wide spreads can be. But using the parameters of the 5 percent rule to determine the appropriateness of spreads across all stocks, without examining the price volatility, trading volume and other characteristics of individual stocks and individual transactions, is exceedingly arbitrary and inaccurate. And arbitrary parameters are no more likely to result in a competitive spread structure than a general command not to post wide spreads. Thus, the excessive mark-up rule should be reserved for its original purpose - targeting individual instances of truly excessive mark-ups - and not for overhauling the spread structure in the Nasdaq market.

Thus, holding market maker liable for their past behavior is more of a bullying tactic than a serious attempt to reform the operation of the market structure. It is like telling a child who returns home from school with a bruised knee that his conduct was bad and punishing him for it, without saying what precisely was bad about his conduct, and, for that matter, without even figuring out what precisely led to the injury. True, the punishment will have a deterring effect, and will likely reduce the probability that the child will bruise his knees in the future. But the deterrence will be very general, holding back the child from participating in a friendly game of soccer in the same way as deterring him from engaging in a school-yard brawl.

VI. REGULATORY PROPOSALS

The discussion so far has shown that there are substantial inefficiencies in the operation of Nasdaq and that these inefficiencies are not appropriately addressed by current legal principles. The recent reduction in spreads should not satisfy regulators that these inefficiencies have been addressed. First, there is no reason to believe that spreads, even after their dramatic decline, are at the competitive level. What's more, since the basic

incentive structure operating on market makers has not changed, there is no reason to believe that, without regulatory reform, spreads will not revert back to their original levels. This section examines what can be done and highlights what should and should not be done to reform Nasdaq.

A. What Not to Do

Ironically, a variety of proposed and some already implemented regulatory reforms comprise a good checklist of what will not work to effectively overhaul the Nasdaq system. These include disclosure of payments for order flow, decimilization of the stock market, and a mechanism to regularly monitor spreads.

1. Disclosure of Payments for Order Flow - The SEC has recently promulgated new rules requiring enhanced disclosure of payments for order flow practices on customer order confirmations, on customer annual account statements and on new accounts.[130] Specifically, the new rules require broker-dealers to disclose on customer confirmations whether they receive payments for order flow. In addition, the new rules require "disclosure of the broker-dealer policies for determining where to route customer orders that are the subject of payments for order flow."[131] The SEC has also proposed a further set of rules for comment that would require disclosure of the per-share and aggregate amounts of payments for order flow.[132]

The rationale behind these regulations seems to be that informing customers about these payoffs would prompt them to shop around for brokers whose order routing practices are not biased by such payments. These "unbiased" brokers would presumably do better in getting their customers the best price. This logic seems to work for exchange-listed stocks which are also traded on the OTC. If transactions routed to the exchange floors have a better opportunity to be executed in between the best bid-ask spread than transactions routed to OTC market makers, and if brokers receive payments for order flow

only for transaction routed to market makers, then an "unbiased" broker is more likely to do better for its customers.

This logic, however, does not hold for Nasdaq securities. There is no evidence to suggest that some market makers are better than others in executing retail trades inside the best-bid ask spread. In fact, the structure of the market suggests that brokers discharge their best execution duty by executing retail trades at the best bid-ask spread. For example, the SOES does not provide any opportunity for retail trades to be executed inside the best bid ask spread. And there is no mechanism for non-automated retail trades to obtain price improvement beyond the inside spread.[133] Payments for order flow, therefore, may have no effect on the price at which retail transactions are executed.

What then is the effect of disclosure? At best, it provides customers with useless information which they may not even be able to decipher. At worst, it will steer sophisticated retail customers in the direction of brokers who obtain the highest payments for order flow. This is because rational customers will understand that they will get the same spread no matter where they place their order, but, at the same time, such customers will expect that brokers who obtain greater payoffs from market makers will be in a better position to reduce commissions and to provide other retail services more effectively. Therefore, customers who are smart enough to understand the mandatory disclosures are also likely to be rational enough to use the information in exactly the opposite way presumed by regulators.

2. ***Decimilization*** - A perennial favorite of commentators, this approach would purportedly reduce the relevance of payments for order flow and, at the same time, narrow spreads.[134] The only drawbacks of this cure-all approach are its "unforeseen consequences," which include, among others, the possibility of making markets too competitive and the market making business not attractive enough, and the possibility of triggering an all out price war that will command many casualties and allow the mammoth

victors to reestablish their primary-exchange monopoly.[135] I am not suggesting that we take these doomsday prophecies seriously, but they do embody an important concern: markets exist in a very sensitive equilibrium and reforms that inflict a big enough shock to the system run the risk of generating unintended results.

Some reforms may be worth taking this risk but decimilization probably isn't; it does nothing to address the skewed incentive structure suggested in this paper. Spreads will continue to be large, as long as market makers have no incentive to improve on them. To a large extent, under the current market structure, market makers behave like monopolists who set spreads at the profit maximizing level, as discussed in Section IV(B)(2)(B). Telling them that they have to quote stocks in increments of 1 cent rather than in eighths will be nothing more than an irritant. If left to their own devices, they will do precisely what they did in the past and quote stocks in some larger increment of their own choice. If under regulatory scrutiny they couldn't get away with doing this, they will quote stocks in the mandated smaller increment. But this will do nothing more than increase their stock monitoring and quote altering costs, as discussed in Section V(B).

This is not to say that decimilization has no useful role. As discussed in Section V(B), the choice of a tick size embodies a trade-off between price accuracy, on the one hand, and monitoring and quote altering costs, on the other. There is no point in reducing the tick size if there are no accuracy benefits to be gained. The question is whether finer increments will allow market makers to better reflect information in their quotes. The answer is probably no if market makers do not regularly make use of the smallest available increment in their current quotes. Therefore, the answer was no before May, 1994, but it is a yes today. The difference between these two periods is the size of the best bid-ask spread. This means that spread width has something to do with how accurately market makers quote their stocks. Thus, before observing the quote setting pattern at the competitive spread width, it is premature to determine whether a smaller tick size is efficient.

This suggests that before playing around with tick sizes, regulators should do something about the fundamental forces that affect spread sizes. Once they do this, regulators can experiment with different tick sizes to determine the most efficient level, keeping in mind the costs associated with smaller tick sizes.[136]

3. Monitoring Mechanism - The NASD is apparently considering using the market's computer system to regularly monitor spreads, for the purpose of putting pressure on dealers to keep spreads narrower.[137] It is not clear what this process would precisely entail, but any regulatory involvement in rate setting seems risky. It is also ironic that a regulatory rate-setting mechanism would have to be implemented in a market that has been configured to stimulate free enterprise, with numerous market makers competing relentlessly for the opportunity of executing transactions. Implementing such a mechanism would be a direct admission of the failure of the Nasdaq concept.

Symbolism aside, a monitoring mechanism would do little to address market inefficiencies. A loose monitoring mechanism that triggers a regulatory response when spreads seem to grow too wide would operate as a general deterrent, no better than the general threat of imposing liability for wide spreads, discussed in Section VI(C). A more intricate system that actually engages in the practice of determining spreads would seem to transfer too much authority to NASD bureaucrats. There is no reason to expect officials who are insulated from the risks and responsibilities of market making to be successful in dictating an efficient spread structure. Meanwhile, market makers would be justifiably bemoaning the loss of their autonomy.

B. What to Do - Recommendations

I. Eliminate the Source of the Problem - This paper has been unequivocal in highlighting the source of Nasdaq' s problems: the rampant practice of making best execution guarantees and the rigid structure of payment for order flow arrangements that it inspires. Eliminating the source of the problem would seem to be a promising way to try and address the problem. Prohibiting market makers from extending best execution guarantees to brokers would restore, at least in theory, the basic incentives for market makers to compete on spreads. It would mean that market makers would obtain no retail transaction volume if they are not on the inside spread, which, aside from prompting market maker to improve their spreads, would discharge their duty, under best execution guarantees, to accept transaction volume at all times and on both side of the spread. This should decrease substantially the costs of market making and lead to an even greater reduction in spreads.

What's more, eliminating best execution guarantees would seem to largely reduce the importance of payments for order flow arrangements. The practice of paying for order flow is made possible by two side effecs of best execution guarantees: the excess spread, which provides market maker a pot of money from which to pay brokers, and other structural support, which makes the practice of routing a broker's entire order flow to one market maker logistically feasible. As both these conditions disappear, payments for order flow will decrease, vanishing entirely with the proper choice of tick-size. Without payments for order flow, the inefficiencies inherent in the two tiered pricing structure will disappear, allowing spreads to narrow to the fully competitive level.

The prohibition against best execution guarantees should be configured as a requirement, imposed upon brokers, to direct their order flow only to market makers whose posted quote matches the best available quote. So brokers would be allowed to route customer sell orders only to a market maker who post the best available bid. Analogously, customer buy orders would have to be directed to a market makers who

posts the best available ask. The market maker, of course, should be free to improve the price and execute the transaction somewhere in between the inside spread.

There are two difficulties with this recommendation: acceptance and potential disruption. It will be hard to convince regulators and market participants that a practice as good natured as extending an offer to honor the best available price is as harmful as this article suggests. Moreover, market makers are likely to lobby aggressively against prohibiting the practice which will make it even harder for regulators to act. The suggested reform may impose substantial disruption in the operation of the market because it fundamentally changes the order routing procedures. Such a fundamental change will have unanticipated consequences, but this is a risk worth taking, considering the potential payoff.

2 Tinker with Limit Order Handling -

A more conservative approach is to change the limit order handling procedures so as to allow natural market forces to operate on spreads. Under this approach, market makers would be restricted from trading at prices that are either equal or more favorable than outstanding limit orders.[138] At the same time, market makers would be required to match equally priced limit orders for execution. This would benefit investors in two ways. First, they will have significant opportunity for price improvement from the potential matching with other limit orders for execution inside the prevailing spread. And, second, by placing limit orders, investors will be able to exert pressure on market makers to narrow quotes, because otherwise market makers' trading behavior will be restricted.

But changing the limit order handling procedures will narrow spreads in an indirect way; it will not address the skewed incentive structure that arises from the operation of best execution guarantees. Moreover, the effectiveness of the approach depends on the willingness of retail investors to place limit orders that will exert the desired pressure on

quotes. So while spreads will likely narrow, there is no guarantee that they will fall to the fully competitive level.

VII. CONCLUSIONS

It is not often that academic literature has a direct and immediate impact on the operation of a market. For a single academic study to generate an overnight drop in prices and a year long controversy, leading to proposals that are likely to fundamentally reform a market, is even less common. This paper suggests yet an even less common possibility, that the conclusions of the study that sparked the entire controversy were misguided.

Rather than relying on an implausible collusion between market makers to explain the odd pattern of Nasdaq quotes, this paper finds the explanation in the skewed structural incentives that are inherent to the Nasdaq system. Best execution guarantees and payments for order flow lead to a bifurcated pricing structure, comprised of spreads and payments for order flow, which significantly disadvantages investors. In this system, investors are over-charged in the spread arena, via a spread structure that can be likened to a monopoly pricing regime. Investors are compensated for these over-charges by market makers' payments for order flow. But it is this paper's contention that payments for order flow cannot efficiently offset the excess spread charged to investors because of timing differentials and differences in the form of the two sets of payments. This leads to a reduction in investor welfare and a net social loss.

The paper's objective is not merely to explain an odd pattern of quotes, but, more importantly, to uncover the set of practices that lead to the structural inefficiencies, to highlight their social costs and to propose effective means of reform. The inevitable conclusion is that regulators should drop their antitrust investigations and instead focus their efforts on reforming the market.

[1] Division of Mkt. Regulation, SEC, *Market 2000: An Examination of Current Equity Market Developments*, at II-11 (1994).

[2] *Id.*.

[3] Scot J. Paltrow, *Inside Nasdaq: Questions about America's Busiest Stock Market*, (First in a six-part series), L.A. Times, Oct. 20, 1994, at A1.

[4] *Id.*

[5] *1994 Nasdaq Fact Book and Company Directory*, at 14.

[6] *The Perils of Payment for Order Flow*, 107 Harv. L. Rev. at 1675 (1994).

[7] Yakov Amihud, Thomas S. Y. Ho, and Rogber A. Schwartz, *Market Making and the Changing Structure of the Securities Industries* 44 (1985).

[8] *Id.*

[9] *Id.*, at 3.

[10] *Id.*, at 4.

[11] *See* 58 Fed. Reg. 52,934, 52,937 (to be codified at 17 C.F.R. pt. 240) (proposed Oct. 13, 1993).

[12] *See Market 2000*, supra, note 1, at V-1.

[13] NASD Manual, Rules of Fair Practice, III-1. Jan, 1995.

[14] NASD Manual, Rules of Fair Practice, Interpretation of the Board of Governors - Execution of Retail Transactions in the Over the Counter Market, A-1.

[15] *Id.*

[16] *See* Paltrow, *supra* note 3.

[17] *See* SEC Proposed Rule Making, 1993 WL 403286 (to be codified at 17 C.F.R. pt. 240).

[18] *See* John C. Coffee Jr., *Brokers and Bribery*, N.Y.L.J., Sep 27, 1990, at 1.

[19] See Payment for Order-Flow, *supra* note 6, at 1676.

[20] Securities Exchange Act Release No. 34902, 1994 WL 587790 (S.E.C.) at 1-2, Oct 27, 1994.

21 *Id.*

22 *Id.*

23 *See* Paltrow, *supra* note 3.

24 NASD officials have been tight lipped, and have refused to disclose what portion of SEOS volume is preferenced, even though they do track the information.

25 Securities and Exchange Act, Payment for Order Flow, Release No, 34-34902 (final rule) (to be codified in 17 C.F.R. 240).

26 NASD Notices to Members, SEC Approves New Small Order Execution System Rules, 94-1.

27 NASD Notices to Members, NTM 88-61 (1988 NASD Lexis 180) Aug 25, 1988.

28 NASD Notices to Members, NTM 91-67, Oct 16, 1991.

29 *Id.*

30 NASD Manual, Rules of Fair Practice and Procedures for The SEOS, a-9, Feb, 1995.

31 *Id.*

32 *Id.*, at c-3(B).

33 *See*, Market 2000, *supra,* note 1, at IV-7

34 *Id.*

35 *Id.*

36 The results of the study by William G. Christie and Paul H. Schultz became public on May 27, 1994. Geffrey Tayor and Warren Gentler, *U.S. Examines Alleged Price Fixing on Nasdaq*, Wall St. J., Oct. 20, 1994, at C1. The study was not published, however, until December. 1994. William G. Christie & Paul H. Schultz, *Why Do Nasdaq Market Markers Avoid Odd-Eighth Quotes?*, 49 J. of Finance 1813 (1994).

37 Taylor & Gentler, *supra* note 36.

38 William Power & Jeffrey Taylor, *U.S. Launches Massive Nasdaq Trading Probe*, Wall St. J., Dec. 7, 1994, at C1.

39 *Id.*

40 Richard Taylor & Warren Getler, *Nasdaq Market Makers Face Review of Practices by U.S. Stock Regulator*, Wall St. J. Europe, Nov. 16, 1994, at 13.

41 Warren Getler, William Power & Jeffrey Taylor, *Nasdaq Head Confronts Price-Collusion Allegations*, Wall St. J., Jan 12, 1995, at C1.

42 Warren Getler & William Power, *Street Insiders Set to Review Nasdaq Market*, Wall St. J., Nov. 21 1994, at C1.

43 *Id.*

44 *See*, Christie & Schultz, *supra* note 36, at 1824.

45 *Id.*, at 1823.

46 *Id.*, at 1820.

47 *Id.*, at 1819.

48 *Id.*, at 1834.

49 *Id.*, at 1835.

50 *Id.*, at 1839.

51 *See* Paltrow, *supra* note 3.

52 *Id.*

53 Scot J. Paltrow, *Nasdaq Spreads Have Narrowed, Study Shows*, L.A. Times, Jan. 5 1995, at D1.

54 *Id.*

55 William G. Christie, Jeffrey H. Harris & Paul H. Schultz, *Why Did Nasdaq Market Makers Stop Avoiding Odd Eighth Quotes?*, 49 J. of Finance 1841 (1994).

56 *Id.*, at 1858.

57 *Id.*, at 1841,1852.

58 *Id.*

59 *Id.*, at 1859.

60 *See* Paltrow, *supra* note 3.

61 *Id.*

62 *See* K. C. Chan, William G. Christie & Paul H. Schultz, *Market Structure and the Intraday Pattern of Bid-Ask Spreads for Nasdaq Securities*, 68 J. of Business 35, 57 (1995). In the study, 42 percent of the market makers in one sample were inactive and 36% of a second sample were considered inactive. *Id.*

63 NASD Manual, Schedule D - XII(2)(a).

64 Scot Paltrow, *Pros Say Many Nasdaq Trades Reported Late*, L.A.Times, Oct. 24, 1994 at D1.

65 NASD Manual, Schedule D - V(2)(a).

66 *See* Paltrow, *supra* note 3.

67 *Id.*

68 *See* Market 2000, *supra,* note 1 at V-5.

69 *See id.*

70 *See id.*

71 E.F. Hutton, [1988-1989 Transfer Binder] Fed. Sec. L. Rep. ¶ 84,303.

72 *See* Market 2000, *supra*, note xx 1 V-6.

73 *See* Christie, *supra* note 36, at 1834.

74 *Id.*, at 1838.

75 *Id.*, at 1822.

76 *Id.*, at 1835.

77 *Id.*, at 1839.

78 *See* Amihud, *supra* note 7, at 44.

79 *Id.*, at 45.

80 *Id.*

81 *Id.*

82 *Id.*

83 *Id.*, at 48.

84 Critics of the practice argue that the price discover function of the market suffers because spreads no longer reflect the true price in which market makers are willing to transact in the stock. I will address this issue in Section VI(B).

85 This is an over simplification because payments for order flow are generally pre-determined in relation to spread setting. This phenomenon will be discussed in Section V(B)(2)(b)(ii).

86 Payments for order flow will still have some significane because the level of spreads shifts in increments of one-eighth, wheras order flow paymenst can be varied in smaller increments. Thus, market makers would pay for order flow when they are not ready to narrow their spreads by a full one-eighth.

87 A recent paper which examines the effects of price matching guarantees also observes anti-competitive effects. Aaron S. Edlin, *Do "Guaranteed Lowest Prices" Guarantee High Prices? How Price Matching Challenges Antitrust?* (Unpublished) ((John M. Olin Working Paper 93-7, University of California at Berkeley). The Edlin paper, however, conditions such anti-competitive effects on some consumers being uniformed. *Id.*, at 5. According to the paper, the disparities in consumer information allow sellers to price discriminate through price-matching policies. The effects of the practice, however, tends to raise prices to all consumers, resulting in sustainable monopoly prices. *Id.*, at 7-14. Best execution guarantees, on the other hand, lead to uncompetitive spreads even though all brokers are fully informed. Best execution guarantees and the rigid Nasdaq order-routing arrangements have a lock-in effect which leads to monopoly prices even though all brokers are fully informed.

88 That some order flow is not governed by best execution guarantees does not mean that it is executed at unfavorable prices. All it means is that in order for such order flow to be executed at the inside spread it has to be routed to market makers whose quotes match the inside spread.

89 Much of the institutional volume is not covered by best execution guarantees. But this volume is generally negotiated and executed inside the best bid ask spread, and therefore has little impact on the posted spreads.

90 *See* Chan, Christie & Schultz, *supra* note 62, at 56.

91 Arguably, these allocational inefficiencies are no worse than similar inefficiencies in a variety of other markets. For example, retail customers who frequent shopping malls at off hours are subsidizing those who shop at peak hours, because both groups are presumably charged the same price while retailers' costs are higher at peak hours. This is true. The only difference is that in the stock trading context we already have a mechanism that can tailor the price to individual transactions, and thereby eliminate allocational inefficiencies. It seems particularly illogical to reintroduce such allocational inefficiencies through a two-tiered pricing structure. Presumably, if retailers incurred the costs of affixing electronic price labels on products, so as to gain the flexibility of adjusting the price on the basis of the cost to service each customer, they would make use of such a capability.

92 It can be argued that this is merely a contractual problem which can be remedied by reconfiguring the contracts for payments for order flow. But the only way to really solve the problem is by determining the payment for order flow at the time the transaction is made, which will allow market makers to fully offset the excess spread by paying for order flow, without assuming any risk from timing differentials. This can be done by telling brokers that their payments for order flow will be determined on the basis of the actual spreads in which their order flow was transacted. But in reality this is no more than a risk shifting mechanism; instead of market makers bearing the risk of the excess spread collapsing, it is brokers who assume the risk that they may not get paid for their order flow. And it would appear that market makers are better positioned to assume this risk because they are the ones determining the spreads on an ongoing basis. In addition, making such trade by trade estimates of the excess spreads will be administratively costly and will inevitably lead to disputes between brokers and market makers as to how to determine such excess spread. Even if market makers are able to shift the risk to brokers by tailoring payments for order flow for individual trades, there is still a question of how brokers would be able to shift this risk to consumers. Short of having variable commission rates that are

determined on the basis of the payments for order flow, this will not be possible, and therefore brokers will charge higher commissions so as to be compensated for the additional risk that they assume from guaranteeing a set commission rate but not being guaranteed a set payment for order flow. Finally, one might ask why aren't consumer better off from market makers' willingness to assume the risk inherent in paying pre-determined rates for order flow. The simple answer is that investors are continuing to bear the risk because they have to pay for the excess spread, no matter how wide it is.

93 Payment for order flow arrangements may reduce the routing costs of customer orders, because they allow brokers to route their order flow to one market maker which will tend to reduce search costs and may provide for more efficient paper routing.

94 *See* Chan, Christie & Schultz, *supra* note 62, at 56.

95 *Id.*, at 58.

96 Notice that best execution guarantees are not beneficial even at times of crisis. The guarantees are meaningful when there is at least one market maker who posts a quote in a particular stock. This may lead to the wrong impression that during market crashes best-execution guarantees would tend to increase liquidity because, as long as one market makers posts a price in the security, all market makers will have to transact at this price. Best execution guarantees would, in a sense, amplify market makers' willingness to buy, which should make it easier for investors to unload their positions. So, when one maverick market maker is willing to buy, say, 1,000 shares of stock at $10 per share, all other market makers will have to honor this price, allowing investors to unload much more than 1,000 shares. Unfortunately, this effect is misleading. In order to avoid having to match the $10 per share price, other market makers, knowing that the maverick is only willing to buy 1,000 shares, will direct 1,000 shares worth of trades to the maverick, forcing him to adjust his quotes or to risk getting stuck with much more than the desired 1,000 shares. When the maverick market maker adjusts his price, the stock price will continue to spiral downwards. And,

in reality, no more than the 1,000 share will have been purchased from investors at the $10 per share price.

97 When placing a limit order, customers take a risk that their transaction will not be executed and that they may later have to seek execution at a price less favorable than that prevalent at the time they place the limit order. The greater the iniside spread and consequent opportunity for price imporvement, the more likely customers are to take the risk of placing a limit order.

98 *Id.*, at 57, Table 4.

99 *Id.*

100 NASD Manual, Schedule D -V(2)(c).

101 *See id.*

102 *See id*, at D-V(2)(d).

103 *See Id.*

104 *See* Paltrow, *supra* note 3.

105 1994 Nasdaq Fact Book & Company Directory, p14.

106 *Id.*

107 *See* Christie, Harris & Schultz, *supra* note 55, at 1855.

108 The analysis in this section focuses on whether, under the behavior described in the model, market makers can be found liable for maintaining a tacit collusion to widen spreads. An alternative approach is to focus on the practice of extending best execution guarantees, and to predicate liability on the basis of the anti-competitive effects of the practice itself, without regard to other behavioral patterns of market makers. Such an approach has been proposed to tackle price matching policies. *See* Edlin, *supra* note 87, at 17-36. Under the proposal, price matching policies can be targeted on the basis of the vertical agreement between buyer and seller, because of its anti-competitive effects. *Id.*, at 31. Even if such liability was plausible under current law, it may not be sensible to apply in the case of market makers. The practice of making best execution guarantees is at a minimum sanctioned by the NASD, and, in some instances, actually sponsored

by NASD regulations. For example, the regulations with respect to SOES set up the system for extending best execution guarantees (or establishing "preferencing" arrangements) in that context. It seems counter-intuitive to hold market makers liable for a practice that is set-up by NASD regulations. And it is doubtful that the justice department has any appetite for holding the NASD liable for violation of the antitrust laws.

109 *See* Phillip Areeda & Louis Kaplow, Antitrust Analysis, at 289.

110 *See* Interstate Circuit v. U.S., 306 U.S. 208 (finding an antitrust violations in the actions of eight movie distributors who, aware of a letter by a movie exhibitor encouraging the imposition of minimum prices, engaged in setting such prices) (1939).

111 American Tobacco v. U.S., 328 U.S. 781 (holding that parallel price increases by several tobacco manufacturer, even as prices of inputs and consumer demand was falling, constituted an antitrust violation) (1946).

112 Theatre Enterprises v. Paramount Film Distributing Corp., 346 U.S. 537 (refusing to find an antitrust violation in different distributors' independent but parallel refusal to grant suburban movie theaters first-run film rights) (1954).

113 *See* Areeda and Kaplow, *supra* note 109, at 307-8.

114 *Id.*, at 312-13.

115 *See also*, Milgram v. Loew's. 192 F.2d 579, 583 (3rd Cir. 1951) cert. denied, 343 U.S. 929 (1952) (holding that each distributor's refusal to license first-run movies to a drive-in theater, even where higher rental was offered, was "in apparent contradiction to its own self-interest," which strengthens the case for a conspiracy); Ambook Enterprises v. Time, 612 F.2d 604 (2d Cir. 1979) (affirming liability of newspapers who imposed different rate structures as to direct advertisers as compared with advertising agencies because the uniform conduct could not be explained by defendants); Viking Theatre Corp. v. Paramount Film Dist. Corp., 320 F.2d 285, 299 (3rd Cir. 1963)("proof of a conspiracy may not rest on similarity of conduct in the absence of evidence that

the alleged wrongdoers were mutually aware of such conduct and that the mutual wareness entered into their decisional processes").

116 *See* Theatre Enterprise, *supra,* note 112.

117 *See* Areeda and Kaplow, *supra* note 109, at 311.

118 *Id.*

119 North Penn Oil & Tire Co. v. Phillips Petro. Co., 358 F.Supp. 908, 923 (E.D.Penn. 1973).

120 *See* Christie, Harris and Schultz, *supra* note 55, at 1854.

121 *See* Areeda and Kaplow, *supra*, note 109, at 308.

122 *See* Ball v. Paramount Pictures, 169 F.2d 317, 319(3rd. Cir. 1948)(inferring conspiracy is proper "when the concert of action 'could not possibly be sheer coincidence'").

123 *See* David L Ratner & Thomas L. Hazen, *Securities Regulation* (4th ed. 1991) at 844-46.

124 *Id.*

125 NASD Manual, Rules of Fair Practice, Art III Sec. 4, Interpretation of the Board of Governors.

126 *See* Alstead v. Dempsey & Co., Sec.Ex.Act Rel. No. 20825, CCH Fed.Sec.L.Rep. ¶ 83,607 (1984).

127 *Id.*

128 *Id.*

129 See Universal Heritage Investments Corp., SEC Release No. 34-19308 (1982) (holdling that mark-ups ranging from 10.1 to 20 percent, calculated on the basis of a market maker's contemporaneous costs in transactions with other dealers, violated NASD's Rules of Fair Practice); A. Bennett Johnson, 45 SEC 278 (1973) (holding that mark-ups ranging from 8.5 to 25 percent, computed in reference to prices paid by the firm in contemporaneous purchases from another dealer, violated NASD's Rules of Fair Practice).

130 Securities Exchange Act Release No. 34902, 1994 WL 587790 (S.E.C.) at 1-2, Oct 27, 1994.

131 *Id.*

132 *Id.*

133 According to statistics obtained directly from the NASD, in Feruary, 1995, approximatedly 86% of transactions of 500 shares or less, and 85% of the trading volume in such transactions, was executed at the inside spread, with no improvement. Since these numbers include transactions between market makers and limit orders, it seems that the chances of a small order being executed in between the inside spread are at best slim.

134 *See* The Peril of Payment for Order Flow, supra, note 6, at 1689.

135 *See id.*, at 1690.

136 Regulators may want to experiment with individual stocks and draw implications for the rest of the market, so as not to wreak havoc in the overall market. Because different stocks will have different competitive spreads, it is not clear that all stocks will have the same efficiecy maximizing tick size. Regulators will have to determine whether it is sensible to make distinctions on this basis between different stocks.

137 William Power, *Nasdaq Weighs Pre-emptive Changes*, The Wall St. J., Mar. 14, 1995 at C1.

138 Alternatively, market makers could only be restriced from trading at prices that are more favorable than outstanding limit orders, therefore allowing them to trade at prices that are equal to oustanding limit orders. Such a rule would seem to have less teeth becuase it would leave market makers substantial trading freedom, without the need to execute pending limit orders.