

Imperfect Market Monitoring and SOES Trading

Thierry Foucault¹ Ailsa Röell² Patrik Sandås³

Current Draft: November 1998

¹Dept. of Finance, HEC, School of Management, 1 rue de la Liberation, 78351 Jouy en Josas, France, Tel: (33) 1 39 67 94 11, Fax: (33) 1 39 67 70 85, E-mail:foucault@hec.fr

²Department of Economics, Princeton University, Princeton, NJ 08544, Tel 609-258-4000, Fax:609-258-6419, E-mail:aroell@princeton.edu

³Finance Department, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104. Tel: 215-898-1697. Fax: 215-898-6200. E-mail:sandas@wharton.upenn.edu

Abstract

We develop a theoretical model of price formation in a competitive dealership market. The model is designed to match some of the key institutional features of the Nasdaq's Small Order Execution System (SOES). Market makers post firm quotes and choose how intensively to monitor the arrival of news. Monitoring is costly and therefore is not optimal for market makers to monitor the news arrival continuously. Imperfect monitoring creates profit opportunities for speculators (who compete with SOES bandits), who also monitor the news arrival in order to pick off 'stale' quotes. The fact that market monitoring is a *public good* for the market makers provides further profit opportunities for the speculators. The presence of speculators widens the spread as market makers protect themselves against the risk of being picked off. Prices are more likely to reflect news when speculators enter the market as the aggregate level of monitoring increases. Thus, in the model the presence of speculators affects the trade-off between the bid-ask spread and price discovery.

1 Introduction

Nasdaq's Small Order Execution System (SOES) allows brokerage firms to automatically execute small orders at the best quotes posted by NASDAQ dealers. The SOES was introduced in 1984 to allow retail investors to quickly execute small orders at posted quotes.¹ Although it was intended for small retail customers, SOES mainly attracted professional day traders (labeled SOES bandits by market makers), who account for a large proportion of the SOES trading volume. The role of the SOES and its alleged impact on Nasdaq trading costs, liquidity, and volatility has been the subject of a long and heated policy debate.² Harris and Schultz (1998) offer direct empirical evidence on the trading profits of SOES bandits. They show that bandits on average make small positive trading profits that are statistically significant. The bandits in their sample made their profits by following momentum strategies and occasionally by trading on public announcements. The positive profits of the SOES bandits are puzzling since they have less information than the market makers.³ The authors suggest that

The existence and profitability of SOES bandits raise new questions about the efficiency of different market structures. Bandits do not have any more information than the market makers that they trade against and in many cases they have less information. But bandits still make money. [...]. We believe the answer is that market makers are inherently less efficient at price discovery than are bandits. [...] bandits have a much greater incentives to concentrate on what they are doing, to follow stock prices closely, and to stay in front of their terminals than do market maker employees. Unusually fast or skillful traders may find SOES trading to be more profitable than working for a Nasdaq market maker.

¹After the crash of 1987 participation in SOES was made mandatory for all market makers. The maximum order size in the SOES has varied over this period between 100 and 1000 shares. See Harris and Schultz (1997) or Battalio, Hatch, and Jennings (1997) for a detailed description of SOES and the strategies of the professional day traders or the so called 'SOES bandits'.

²Many argue that the SOES bandits increase trading costs and decrease liquidity in the market. In a Washington Post article, on February 7, 1994, Joseph Hardiman, president of the National Securities Dealers Association, said that 'The SOES activists were picking off market makers, who were slow to adjust. The losses to SOES activists made market makers gun shy, causing them to widen their price spreads.' Opponents argue that the SOES bandits improve the price discovery process. In a testimony before the House Committee on Commerce, David Whitcomb, said that 'Abolishing SOES would remove the 'market discipline', which keeps market makers on 'their toes' and causes prices to rapidly adjust when news occurs.'

³The SOES bandits also pay commissions on their trades.

In this paper, we develop a theoretical model of imperfect monitoring by market makers and show how this phenomenon affects price formation. In our model, market makers post firm quotes and select how intensively they monitor the arrival of news. Continuous monitoring of news is costly and therefore market makers never monitor news continuously in equilibrium. Less than perfect monitoring of news by the market makers create occasional profit opportunities due to ‘stale’ quotes. A second group of agents referred to as speculators, also monitor the news arrival but do not post quotes. The speculators exploit the profit opportunities created by imperfect monitoring by trading with the market makers. The market makers losses from trading with the speculators are offset by gains from trading with liquidity traders. The behavior of the speculators is similar to the behavior of the SOES bandits described in Harris and Schultz (1998). Our model provides an explanation for the puzzling stylized fact reported in Harris and Schultz (1998), which does not depend on differences in monitoring costs, differences in ability, or agency problems in trading.⁴

Our main results are:

- Imperfect market monitoring due to monitoring costs will create profit opportunities or free trading options. These opportunities allow speculators, who only trade on public information, to recover their monitoring costs by ‘picking off’ exposed quotes.
- We show that the monitoring activity is a public good for the market makers. Market makers can free ride on other market makers’ monitoring efforts and hence all market makers invest less in monitoring than they would in the absence of this externality. This further increases the profit opportunities for the speculators.
- Within our model the presence of the speculators increases the bid-ask spread and increase the likelihood that prices reflect new information.
- Our results suggest that an increase (decrease) in the maximum size for which dealers’ quotes are binding lead to entry (exit) of more speculators and a higher (lower) bid-ask spread. The increase (decrease) in the spread is accompanied with a increase (decrease) in the probability that prices reflect news.
- Positive market maker profits are possible in equilibrium even if market makers compete *à la Bertrand*. This is due to the fact that a market maker who post an

⁴However, our model can be extended to consider possible agency problems in trading. The monitoring decision can be interpreted as an unobservable (and non-contractable) choice of effort made by the trader, who is an employee at a market making firm. It is also straightforward to analyze the effects of differences in skills or monitoring costs in our setting.

aggressive quote (undercut) will have to bear more of the monitoring costs (other market makers will have a greater incentive to free-ride).

Our model is most closely related to Copeland and Galai (1983), who analyze the free-trading option aspect of fixed quotes. Our model extends their analysis by introducing a monitoring cost and by explicitly analyzing the monitoring decisions of market makers and speculators. Several authors, for example, Kleidon and Willig (1995), argue that the odd-eighth avoidance, documented by Christie and Schultz (1994), could be a rational response by the dealers facing additional costs of making a market in a stock. Kandel and Marx (1998) develop a theoretical model to study whether odd-eighth avoidance is a rational response by the dealers to SOES bandits trading. In their model the profit opportunities of the SOES activists are implicitly assumed to be due to imperfect monitoring by the dealers. Their main goal is to study the interaction between the tick size, odd-eighth avoidance, and SOES trading. Our focus is different. We are primarily interested in analyzing how profitable SOES trading can occur in a competitive market. Our contribution is to show how the SOES profit opportunities arise in equilibrium in a market where it is costly to monitor the flow of information. Other differences between Kandel and Marx (1998) and our paper are that we determine the equilibrium bid-ask spread (the spread is exogenous in their paper), and we explicitly show how profit opportunities arise in equilibrium as a result of imperfect monitoring decisions by the market makers. This is an advantage because it allows us to study how changes in exogenous variables, e.g., the maximum SOES order size, affect the bid-ask spread, the monitoring decisions, and the entry decisions of speculators. This is important since many variables, e.g., trading activity and volatility, are believed to affect the bid-ask spreads as well as the profits of the SOES bandits.

Several recent empirical studies of SOES trading on the Nasdaq are also important for our paper. Battalio, Hatch, and Jennings (1997), using vector autoregressions, show that SOES bandits may intensify the price discovery process. Harris and Schultz (1997) analyze trading in the 20 largest Nasdaq stock around a rule change that decreased the maximum SOES trade size from 1000 to 500 shares. They conclude, based on vector autoregressions, that maximum size SOES trades are more likely to be motivated by information than other trades. Furthermore, they document a substantial shift in the frequency of 1000-share and 500-share trades at the time of the rule change.⁵ Harris and Schultz (1997) also provide evidence suggesting that the rule change had its intended effect of reducing the marker

⁵Barclay, Christie, Harris, Kandel, and Schultz (1998) document similar shifts in a study of the effects of the SEC order handling rules introduced in 1997.

makers' losses to SOES bandits. Benston and Wood (1998) use a measure of the number of maximum-sized trades submitted to SOES to measure the level of the SOES bandits' activity. Secondly, they compute measures of the adverse selection costs incurred by market makers for different order sizes. By combining these two measures and taking advantage of the changes in the SOES usage and rules over a ten-year sample period they quantify the effects of mandatory SOES on trading costs. We contribute to this literature by showing that the presence of speculators (SOES bandits) can improve the price discovery process even when all market participants have access to the same information and face the same costs. In our model, a decrease in the maximum order size for which the market makers' quotes are binding lead to the entry of fewer speculators. Our model can also generate new testable predictions by explicitly showing how costs of monitoring and stock specific characteristics affect the profitability of trading by SOES bandits.

The rest of the paper is organized as follows. The general features of the model are presented in the next section. In section 3 the equilibrium with a fixed number of speculator is analyzed. In section 4 we endogenize the number of speculators. In Section 5 we briefly discuss how the model relates to the empirical literature. The final section concludes. All proofs are in the appendix.

2 The Model

In this section, we describe the general features of the trading game that we use for analysis in the later sections.

2.1 Timing and Traders

We consider the market for a risky asset. The liquidation value of the risky asset is uncertain and is denoted $\tilde{V}$. There are 3 types of traders in this market: (i) M *dealers* who post quotes and monitor the market, (ii) N *speculators* who monitor the market and submit market orders when they perceive profit opportunities and (iii) a liquidity trader. $\mathcal{M}$ and $\mathcal{N}$ will denote the set of all dealers and all speculators, respectively. The expected value of the asset at the beginning of the trading day is denoted v_0 . We normalize the asset expected value to zero ($v_0 = 0$).

The trading day comprises 3 stages. In the first stage, *the quoting stage*, the risk neutral dealers quote *simultaneously* an ask price at which they offer to sell shares of the risky

asset.⁶ The market organization is such that the dealers are required to honor their quotes for up to Q shares of the risky asset and that execution is automatic for all order sizes lower than Q . For orders larger than Q , dealers can back away from their quote or trade at their posted offer. We denote $A = (a_1, \dots, a_i, \dots, a_M)$, the vector of ask prices posted by the dealers in stage 0. Since a_i is also the difference between the ask price and the asset expected value (because $v_0 = 0$), we often refer to a_i as the *spread* posted by dealer i . In the second stage, *after* observing the quotes posted in the market, the dealers and the speculators choose their monitoring level. The monitoring level chosen by a trader determines the probability that he is the first to react in the event of a public announcement regarding the asset value. The monitoring decision is described in more details below. In the third stage, one of the three following events occurs. With probability $\alpha < 1$, information arrives indicating an increase in the asset value. In this case, the new expected value becomes: $v_1 = +\sigma$. With probability $(1 - \alpha)\beta > 0$, a buy market order is submitted by a liquidity trader. The expected size of the liquidity trader's order⁷ is δQ . Finally with probability $(1 - \alpha)(1 - \beta)$, nothing happens. In the case of new information, the first trader who reacts to the new information is denoted f . If f is a dealer then he updates his quote. If f is a speculator then she can submit a market buy order. A market order is always executed against the best quote in the market. If $M \leq N$ dealers are tied at the best price, each dealer will buy or sell a fraction $1/M$ of the order.

Our assumptions concerning the market structure closely reflects the organization of trading in the Nasdaq's SOES trading system. Q is the maximum order size for which automatic execution is possible. This maximum order size in NASDAQ has changed several times. It has been reduced from 1000 shares to 500 shares in January 1994, for most stocks, on a trial basis. The maximum order size returned to 1000 shares in March 1995. With the implementation of new order handling rules in 1997, it is now 100 shares. The speculators can be thought of as the SOES bandits. A speculator cannot trade more than one time against the dealers. This assumption is consistent with NASDAQ rules which prohibit SOES traders from initiating more than one position in the same stocks within 5 minutes. Another feature of Nasdaq is that dealers execute, at their posted offers, orders that are larger than the maximum size for which automatic execution is possible. SOES bandits typically do not take part to these trades since they are negotiated by phone. This slows

⁶We consider only the determination of the sell prices. Symmetric results can be obtained for bid prices.

⁷As usual in the market microstructure literature, the motives for trading of the liquidity trader are not modelled.

down the execution process and dealers can back away from their quote upon realizing that the counter-party is a bandit (See Harris and Schultz (1997)). This suggests that, in our model, there is no reason to restrict our attention to values of δ that are lower than 1.

2.2 Market Monitoring

We model market monitoring by the speculators and the dealers as follows. Let $\lambda_i \in [0, +\infty)$ be the *monitoring level* of market-maker $i \in \mathcal{M}$ and let $\gamma_j \in [0, +\infty)$ be the monitoring level of speculator $j \in \mathcal{N}$. If there is a new information indicating an increase in the asset value, we denote $P(\lambda_i)$ the probability that dealer i is the first trader to observe this innovation. In the same way we denote $P(\gamma_j)$ the probability that speculator j is the first trader to observe the innovation in the asset value. We assume that $P(\lambda_i)$ and $P(\gamma_j)$ are given by:

$$P(\lambda_i) \equiv \text{Prob}(f = i) \equiv \frac{\lambda_i}{\lambda_i + \sum_{\{k \in \mathcal{M}, k \neq i\}} \lambda_k + \sum_{\{j \in \mathcal{N}\}} \gamma_j} \quad \forall i \in \mathcal{M} \quad (1)$$

$$P(\gamma_j) \equiv \text{Prob}(f = j) \equiv \frac{\gamma_j}{\gamma_j + \sum_{\{k \in \mathcal{N}, k \neq j\}} \gamma_k + \sum_{\{i \in \mathcal{M}\}} \lambda_i} \quad \forall j \in \mathcal{N} \quad (2)$$

First note that $P(\cdot)$ increases with the monitoring level chosen by a trader. This captures the intuition that the more closely a trader monitors the market, the larger the chance that he will be able to react on public announcements before the other traders. Second note that the larger the aggregate monitoring level of the other traders, the lower $P(\cdot)$. This captures the intuition that the larger the monitoring levels of other traders, the lower the probability that a trader will be the first to react to a public announcement. Finally note that $P(0) = 0$ and that $P(+\infty) = 1$. This means that a monitoring level equal to zero corresponds to the decision of *not monitoring the market at all*. On the contrary, an infinite monitoring level corresponds to the decision of *continuously monitoring the market*. For any monitoring levels, in between these two extremes, there is some monitoring but it is imperfect.

Harris and Schultz (1998) point that monitoring is not perfect because it requires some effort. In our setting, the monitoring level chosen by a trader characterizes *the effort* he or she puts in monitoring the market. We denote the monetary disutility associated to a given monitoring level by $\Psi(\cdot)$. We assume:

$$\Psi(\lambda_i) = \frac{c_d}{2} \lambda_i^2 \quad \forall i \in \mathcal{M} \quad (3)$$

and

$$\Psi(\gamma_j) = \frac{c_s}{2} \gamma_j^2 \quad \forall j \in \mathcal{N}. \quad (4)$$

We call c_j , the monitoring cost level for trader j . Let us define $r \equiv \frac{c_d}{c_s}$. A ratio r higher (lower) than 1 indicates that, for a given level of monitoring, the marginal cost of monitoring is higher (lower) for the dealers than for the speculators.

The traders (speculators and dealers) choose their monitoring level *simultaneously* in the second stage of the trading round, after observing the quotes posted in the market. We call this stage, *the monitoring stage*. The monitoring levels will depend on the offers posted in the market. Thus we denote $\lambda(A) = (\lambda_1(A), \dots, \lambda_M(A))$, the vector of monitoring levels chosen by each dealer given the offers posted in the market. In the same way, $\gamma(A)$ is the vector of monitoring levels chosen by each speculator.

In order to protect themselves against the risk of being picked off, dealers can develop market-monitoring programs that alert them when a quote update or a transaction occur in the market.⁸ Accordingly a SOES bandit must react immediately upon observing a quote update or a public announcement if she wants to have a chance to trade before the dealers adjust their quotes. This is emphasized by Harris and Schultz (1998) who find that ‘SOES bandits profit by purchasing before *most* dealers raise their prices’ (Harris and Schultz (1998), p.42). Thus we assume that dealers are picked off if and *only if* the first trader who reacts to the public announcement is a speculator. If the first trader who reacts is a dealer, we assume that the other dealers become automatically aware of his quote update and adjust their offer immediately. It follows that a speculator who is first to react to the arrival of new information submits a buy market order of size Q if the best ask price in the market is lower than the asset expected value conditional on the new information: $v_1 = +\sigma$. If she is not first to react, she does nothing since the dealers immediately update their offers after the trade of another speculator or the quote update of another dealer. The trading game stops when either (i) a transaction occurs or (ii) all the dealers have updated their quotes. Figure 1 depicts the different stages of the trading game. Table 1 provides the payoffs of the different traders.

⁸The Autoquote Policy of Nasdaq bans the automated update of quotations by market makers. However It does not prevent dealers from designing market-monitoring programs that keep track of changes in the market and to manually update their offer when they receive an alert.

2.3 Equilibrium.

We search the subgame perfect equilibria of the trading game, i.e. a set $\{A^*, \lambda^*(A^*), \gamma^*(A^*)\}$ such that $\lambda^*(A)$ and $\gamma^*(A)$ form a Nash equilibrium of the monitoring stage for all possible vectors of quotes, A , chosen in the first stage and such that A^* is a Nash equilibrium of the quoting stage.

In practice, SOES bandits trade on public news announcements.⁹ They also trade on market makers' quote up-dates. In this case, they pick off market makers, who are slow to update their quotes when a trend emerges. In addition bandits may try to spot market maker trading patterns, which suggests that the market makers 'prepare' to execute an institutional trade. In our setting, in order to simplify the analysis, the speculators trade exclusively on public news announcements. Note that we could easily modify the trading game so that the speculators trade on the basis of a quote update rather than a public announcement. This could be done as follows. When a new information arrives, one dealer, randomly selected, becomes aware of the information and updates his offer. Then the first trader who becomes aware of the quote update is determined as explained above and the game proceeds as in the case of a public announcement.

In all the analysis, we treat the number of dealers as exogenous. Recall that our final objective is to use the model in order to derive testable predictions concerning the behavior of prices and the activity of traders in the SOES of Nasdaq. The number of dealers on Nasdaq is likely to be determined by all the trades they execute, not just those going through SOES. Throughout the analysis we will therefore take the number of market makers as given. In contrast, the number of SOES bandits is certainly determined by the profitability of the trades they execute in SOES. Thus, in Section 4, we endogenize the number of speculators.

3 Equilibrium with a Fixed Number of Speculators

In this section, we solve for the subgame perfect equilibria of the trading game taking the number of speculators and dealers as given. We proceed by backward induction.

⁹Houtkin (1998) describes the trading strategies followed by SOES bandits. He provides a list of the events that can trigger trades by the SOES bandits. Examples of these events include: business news and earnings announcements, price movements in related stocks, brokerage firms' upgrades and downgrades of stocks, announcements of various economic indicators. All these events fall in the category of what we call a public news announcement.

3.1 Equilibrium in the monitoring stage

Assume that the best ask price posted in the quoting stage is $a^* < \sigma$ and that all the dealers are posting this offer. Let $\Pi_d(\lambda_i, \lambda_{-i}, \gamma, a^*, M)$ be dealer i 's expected profit when M dealers post an ask price equal to a^* in the quoting stage, for given vectors λ_{-i} and γ of monitoring levels, respectively for the other dealers and for the speculators. Using the payoff table (Table 1), we get:

$$\Pi_d(\lambda_i, \lambda_{-i}, \gamma, a^*, M) = -\alpha \text{Prob}(f \in \mathcal{N}) \frac{Q}{M} (\sigma - a^*) + (1 - \alpha) \beta \frac{\delta Q}{M} a^* - \frac{c_d}{2} \lambda_i^2. \quad (5)$$

$\text{Prob}(f \in \mathcal{N})$ is the probability that dealer i will be picked off, conditional on a new information arrival. Actually recall that we have assumed that a dealer is picked off if and only if the first trader who reacts to a new information is a speculator. Using Equations (1) and (2), we obtain:

$$\text{Prob}(f \in \mathcal{N}) = \text{Prob}(\cup_{i=1}^{i=M} \{f = i\}) = \frac{\sum_{j=1}^{j=N} \gamma_j}{\lambda_i + \sum_{m \neq i}^{m=M} \lambda_m + \sum_{j=1}^{j=N} \gamma_j}. \quad (6)$$

Note that this probability decreases with the monitoring level chosen by dealer i but also with the aggregate monitoring level of the remaining dealers ($\sum_{m \neq i}^{m=M} \lambda_m$). Thus the monitoring activity of each dealer indirectly benefits the other dealers. For this reason, the monitoring activity of each dealer is a *public good*. Actually, when a dealer updates his quote, he signals to the other dealers that their own offers are mis-priced. In such a case, the other dealers update their offers and obtain protection against the risk of being picked off.

Given the monitoring levels chosen by the other traders, dealer i will choose the monitoring level which maximizes $\Pi_d(\lambda_i, \lambda_{-i}, \gamma, a^*, M)$. Thus the monitoring level chosen by dealer i solves:

$$\frac{\partial \Pi_d(\lambda_i, \lambda_{-i}, \gamma, a^*, M)}{\partial \lambda_i} = 0 \quad (7)$$

Using Equation (5), this equation can be written:

$$-\alpha \frac{\sum_{j=1}^{j=N} \gamma_j}{(\lambda_i + \sum_{m \neq i}^{m=M} \lambda_m + \sum_{j=1}^{j=N} \gamma_j)^2} \frac{Q}{M} (\sigma - a^*) - c_d \lambda_i = 0 \quad (8)$$

It is direct that the Second Order Condition is satisfied if $a^* < \sigma$, which will be the

case in equilibrium. Let $\Pi_s(\gamma_j, \lambda, \gamma_{-j}, a^*)$ be the expected profit for speculator j , for given monitoring levels of the other traders.

$$\Pi_s(\gamma_j, \lambda, \gamma_{-j}, a^*, N) = \alpha \text{Prob}(f = j) Q(\sigma - a^*) - \frac{c_s}{2} \gamma_j^2 \quad (9)$$

In case of a change in the asset value, a profit opportunity arises from the fact that dealers' offers are temporarily mispriced. As discussed at the beginning of this section, a speculator can capture this profit opportunity *iff* she is the first to react to the public announcement of a change in the asset value. The probability that speculator j will react first is $\text{Prob}(f = j)$ which is given by Equation (2). Speculator j chooses the monitoring level that maximizes $\Pi_s(\gamma_j, \lambda, \gamma_{-j}, a^*, N)$, i.e., γ_j solves:

$$\alpha Q(\sigma - a^*) \frac{\sum_{m=1}^{m=M} \lambda_m + \sum_{s \neq j}^{s=N} \gamma_s}{(\gamma_j + \sum_{m=1}^{m=M} \lambda_m + \sum_{s \neq j}^{s=N} \gamma_s)^2} - c_s \gamma_j = 0. \quad (10)$$

We solve for the symmetric Nash equilibrium of the monitoring stage, i.e., the equilibrium in which all the dealers choose the same monitoring level λ^* and all the speculators choose the same monitoring level γ^* . These levels must be solution of Equations (8) and (10) for all the dealers and all the speculators. It follows that λ^* and γ^* solve the following system of equations:

$$\alpha \frac{N \gamma^*}{(M \lambda^* + N \gamma^*)^2} \frac{Q}{M} (\sigma - a^*) - c_d \lambda^* = 0, \quad (11)$$

and

$$\alpha \frac{M \lambda^* + (N - 1) \gamma^*}{(M \lambda^* + N \gamma^*)^2} Q(\sigma - a^*) - c_s \gamma^* = 0. \quad (12)$$

We obtain the following result.

Lemma 1 : *If M dealers post the best ask price a^* , the dealers and the speculators choose the following monitoring levels in equilibrium:*

$$\gamma^*(a^*) = \sqrt{\frac{\alpha Q(\sigma - a^*)}{c_s} \frac{\sqrt{(C + (N - 1))}}{C + N}}, \quad (13)$$

$$\lambda^*(a^*) = \frac{C}{M} \gamma^*, \quad (14)$$

where $C \equiv 0.5(\sqrt{(N-1)^2 + \frac{4N}{r}} - (N-1))$.

The monitoring level of both types of traders is dependent on the equilibrium level of the spread. The spread is endogenous and is determined by the parameters of the model. Consequently we first need to solve for the spread in order to study the determinants of the monitoring level.

Equation (14) shows that there is a positive relationship between the monitoring level of the speculators and the monitoring level of the dealers. This means that securities that are closely monitored by the speculators are also closely monitored by the dealers.

In equilibrium, the unconditional probability of being picked off for a dealer is not dependent on the spread posted in the market, however. Actually, using Lemma 1 and Equation (6), we obtain that:

$$\alpha \text{Prob}(f \in \mathcal{N}) = \frac{\alpha N}{C + N}. \quad (15)$$

This probability plays an important role since it determines the adverse selection risk faced by the dealers because of imperfect monitoring. The probability of being picked off for a dealer does not depend on the equilibrium ask price chosen by the dealers for the following reason. When the dealers increase their ask price, speculators monitor the market less intensively. Accordingly the dealers also monitor the market less intensively (see Equation (14)). This concomitant decrease in the monitoring levels of the speculators and the dealers is such that, *in equilibrium*, the probability of being picked off is unaltered by the change in the offers posted in the market. The probability of being picked off is also independent of the number of market makers¹⁰. The intuition is the following. Note that, for a given spread, γ^* does not depend on M . It follows from Equation (14) that when the number of dealers increases, each dealer chooses a lower monitoring level. Actually, each dealer can free ride on the monitoring effort of a larger number of dealers. The aggregate monitoring of all the dealers, $M\lambda^*$, is unchanged, however. This explains why the monitoring level γ^* chosen by each speculator is insensitive to the number of dealers. It follows from these two remarks that the probability of being picked off does not depend on M .

¹⁰This property does not hold anymore when the number of speculators is endogenous because the number of speculators depends on the number of dealers. See Section 4.

3.2 Equilibrium in the quoting stage.

Using Equations (5), (15) and the characterization of monitoring levels, we get:

$$\Pi_d(\lambda^*(a^*), \gamma^*(a^*), a^*, M) = -\alpha \frac{Q}{M} (\sigma - a^*) R(M, N, r) + (1 - \alpha) \beta \frac{\delta Q}{M} a^*, \quad (16)$$

with

$$R(M, N, r) \equiv \frac{N}{C + N} + \left(\frac{rC^2}{2}\right) \frac{(C + N - 1)}{M(C + N)^2}. \quad (17)$$

Using the definition of C , algebra yields: $rC(C + (N - 1)) = N$. Consequently, $R(M, N, r) = \frac{N}{C + N} + \frac{CN}{2M(C + N)^2}$. An ask price equal to a^* is a Nash equilibrium of the quoting stage if (i) no dealer has an incentive to raise his offer given that the other dealers post an offer equal to a^* and (ii) no dealer has an incentive to undercut the equilibrium ask price. The first condition imposes that:

$$\Pi_d(\lambda^*(a^*), \gamma^*(a^*), a^*, M) \geq 0. \quad (18)$$

Let $\hat{a}(r, M, N, \delta)$ be the ask price such that this equation is binding. Using Equation (16), we get:

$$\hat{a}(r, M, N, \delta) = \alpha \sigma \left(\frac{R(M, N, r)}{\alpha R(M, N, r) + (1 - \alpha) \beta \delta} \right). \quad (19)$$

In equilibrium, the ask price must be at least equal to $\hat{a}(M)$ for the dealers to break-even. Now suppose that a dealer undercuts slightly the equilibrium ask price. In this case, the other dealers do not monitor since only the dealer offering the best quote has a chance to be executed. Thus everything is as if there was only one dealer in the monitoring stage. It follows that the expected profit of the dealer who undercuts is $\Pi(\lambda^*(a^*), \gamma^*(a^*), a^*, 1)$. In equilibrium, the dealer must be better off not undercutting (Condition (ii)). This requires:

$$G(a^*) = \Pi(a^*, \lambda^*(a^*), M) - \Pi(a^*, \lambda^*(a^*), 1) \geq 0. \quad (20)$$

Computations yield:

$$G(a^*) = -\frac{(M - 1)}{M} [a^* (\alpha Q \bar{R}(M, N, r) + (1 - \alpha) \beta \delta Q) - \alpha Q \sigma \bar{R}(M, N, r)], \quad (21)$$

where

$$\bar{R}(M, N, r) = \frac{N}{C + N} + \frac{(1 + M)N}{2M} \frac{(C + N - 1)}{(C + N)^2}. \quad (22)$$

Clearly $G(\cdot)$ decreases with a^* . Let $\bar{a}(M)$ be the ask price such that $G(\bar{a}(M)) = 0$. Computations yield:

$$\bar{a}(M) = \alpha\sigma\left(\frac{\bar{R}(M, N, r)}{\alpha\bar{R}(M, N, r) + (1 - \alpha)\beta\delta}\right). \quad (23)$$

In equilibrium, the ask price posted by the dealers must be lower than $\bar{a}(M)$.

The lowest possible equilibrium ask price, $\hat{a}$, increases with the value of $R(r, M, N)$. In the same way, the highest possible equilibrium ask price, $\bar{a}$, increases with the value of $\bar{R}(r, M, N)$. It is straightforward that $R(M, N, r) < \bar{R}(M, N, r) < 1$, for all the values of the parameters. It follows that $\hat{a} < \bar{a} < v_1$. The second inequality implies that a speculator purchases the asset when he is the first trader who observes the new information, as conjectured until this point. Thus we have proved the following result.

Proposition 1 : $A^* = (a^*, a^*, \dots, a^*)$ is a Nash equilibrium of the quoting stage if and only if $a^* \in [\hat{a}(r, M, N, q), \bar{a}(r, M, N, q)]$

Dealers compete *à la Bertrand* and still we obtain that ask prices ($a^* > \bar{a}(r, M, N, \delta)$) such that dealers obtain strictly positive expected profits are possible. The intuition for this result is as follows. When a dealer undercuts his competitors, two effects are at work. On one hand, the dealer captures a larger part of the order flow. On the other hand, the dealer monitors more intensively the market. Actually his competitors lose price priority but, for this reason, face no risk of being picked off. As a consequence, they do not monitor the market at all. It follows that the dealer who undercuts can no longer free ride on the market monitoring efforts made by the other dealers and he therefore must devote more effort to market monitoring. These two effects have opposite impacts on the expected utility of a dealer. The first effect (larger share of the order flow) increases a dealer's trading profit¹¹ but the second effect increases his market monitoring cost. It turns out that for all ask prices below $\bar{a}(r, M, N, \delta)$, the second effect is larger than the first. Dealers are better off not undercutting their competitors and indirectly benefit from their monitoring. Consequently, the positive externality associated with market monitoring by the dealers can result in spreads that are above the competitive level.

The next corollary analyzes the impact of monitoring costs on the spread posted by the dealers.

¹¹ A dealer's trading profit is equal to his expected profit *net* of the monitoring cost. The same definition applies to the speculators.

Corollary 1 : (spread and monitoring cost) *The lower and the upper bounds of the set of possible equilibrium prices have the following properties:*

- $\frac{\partial \hat{a}}{\partial r} > 0$ and $\frac{\partial \bar{a}}{\partial r} > 0$,
- $\lim_{r \rightarrow 0} \hat{a} = \lim_{r \rightarrow 0} \bar{a} = 0$,
- $\lim_{r \rightarrow +\infty} \hat{a} = \lim_{r \rightarrow +\infty} \bar{a} = \frac{\alpha\sigma}{\alpha + (1-\alpha)\beta\delta}$.

The first part of the corollary establishes that the spread increases when the monitoring cost for the dealers increases relative to the monitoring cost for the speculators. The intuition is as follows. Other things equal, when dealers' monitoring cost level increases, λ^* decreases, i.e., dealers put less effort in market monitoring. It follows that the probability of being picked off increases (see Equation (15)). Consequently dealers must increase their ask prices in order to break-even. The second part shows that imperfect market monitoring is the source of the trading costs in our model. Actually when the monitoring cost borne by the dealers disappears, they monitor more and more intensively the market. As a consequence, the risk of being picked off vanishes and the ask prices posted in the market converge to the asset expected value ($v_0 = 0$).

The previous corollary establishes that, in our setting, the spread is due to the monitoring cost borne by the dealers. From now on, we assume that $r = 1$, which entails $C = 1$. This assumption does not affect qualitatively the results that are provided below but simplifies the computations. Furthermore, it makes sure that the testable implications that are derived below do not come from *ad hoc* assumptions on differences in monitoring skills between the dealers and the speculators.

Corollary 2 :

- *An increase in the number of speculators results in an increase in the upper bound and the lower bound of the set of possible equilibrium prices.*
- *An increase in the number of dealers results in a decrease in the upper bound and the lower bound of the set of possible equilibrium prices.*

When the number of speculators increases, dealers choose a lower monitoring level, *ceteris paribus*. Thus, their monitoring cost decreases. But at the same time, the risk of being

picked off increases. Accordingly the expected loss when a dealer is picked off is larger. Overall the cost of market making is larger and dealers must increase their spreads in order to break-even. This explains the first part of the corollary. The inverse relationship between the spread and the number of dealers is well-documented empirically. The literature has provided 2 explanations for this stylised fact. First, competitive pressures are stronger when the number of dealers is larger and second, there is better risk sharing among the dealers. Our model points to a third possible explanation. When the number of dealers is large, dealers can better share the burden of monitoring the market. Each dealer chooses a lower monitoring level but the aggregate monitoring level ($M\lambda^*$) is unchanged. For this reason the probability of being picked off is unaltered by the change in the number of dealers but the monitoring cost for each dealer is now lower. Due to this cost sharing effect, the ask price for which dealers break-even is lower.

Corollary 3 :

- *An increase in δ results in a decrease in the upper bound and the lower bound of the set of possible equilibrium prices.*
- *An increase in β results in a decrease in the upper bound and the lower bound of the set of possible equilibrium prices.*
- *An increase in σ results in an increase in the upper bound and the lower bound of the set of possible equilibrium prices.*

Dealers compensate the losses they incur when they are picked off by speculators by the gains from trading with liquidity traders (see Equation (5)). Speculators always place maximum-sized orders. Consequently, when the average order size of liquidity traders increases, the gain obtained when trading with a liquidity trader increases relative to the loss incurred when trading with a speculator. It follows that dealers post lower spreads. When β increases, the probability of liquidity trading increases, other things equal. Thus dealers decrease their spreads. Finally when σ increases, the loss when a dealer is picked off is larger. Furthermore, for a given spread, dealers monitor more intensively the market when changes in the asset value are larger (see Equation (14)). Thus their market monitoring cost increases. As all components of the cost of market-making increases, dealers raise their spreads. In the same way, we could show that spreads enlarge when α increases. Markets

with (i) a high frequency of new information and/or (ii) large changes in value when information arrives are more volatile. Thus, in our setting, larger spreads are associated to more volatile markets.

Corollaries 2 and 3 are useful in order to develop intuitions for how the model works. They also show that our model has reasonable properties. At this stage, the model yields testable implications regarding the bid-ask spread and the monitoring decisions of the market makers and speculators. However, these predictions may be misleading since the number of speculators is likely to be endogenous and dependent on other parameters that influence the spread, in particular the number of dealers. For the model to be amenable to empirical analysis, we need to endogenize the number of speculators. This is the purpose of the next section.

4 Measures of Speculators Activity

In this section, we study the determinants of the number of speculators. This allows us to identify the determinants of the probability of observing a trade by a speculator. This is of interest since the level of trading by SOES bandits has been the focus of the empirical literature.

4.1 Equilibrium with an endogenous number of speculators

We assume that a trader decides whether to become a speculator or not before the quoting stage of the trading game. A speculator bears a fixed cost K , that is sunk at the entry stage. The number of speculators is then determined in such a way that the expected trading profit of a speculator is just equal to the fixed cost borne by the speculator¹². The expected trading profit for the speculators depends on the ask price posted by the dealers in the quoting stage. For brevity, we just consider the case in which the dealers post the zero-profit ask price, $\hat{a}$, that is the lower bound of all possible equilibrium prices. Qualitatively, similar results can be obtained by considering the other possible equilibria of the quoting stage. Furthermore note that the expected profit of a speculator cannot be larger than $\alpha Q\sigma$. Thus, we assume that $K < \alpha Q\sigma$, otherwise there is no entry by speculators.

¹²As usual in this type of analysis, we ignore the integer problem.

Using Equation (9) and Lemma 1, we obtain that the expected profit of a speculator is given by:

$$\Pi_s(\lambda^*(\hat{a}), \gamma^*(\hat{a}), \hat{a}, N) = (\sigma - \hat{a})Q\alpha\left(\frac{(N+2)}{2(N+1)^2}\right). \quad (24)$$

It has been shown in the previous section that $\hat{a}$ increases with N , other things equal. Thus $\Pi(\lambda^*(\hat{a}), \gamma^*(\hat{a}), \hat{a}, N)$ strictly decreases with N . Using Equation (19), Equation (24) can be rewritten as:

$$\Pi_s(\lambda^*(\hat{a}), \gamma^*(\hat{a}), \hat{a}, N) = \left(\frac{2+N}{2(1+N)^2}\right)\left(\frac{\alpha(1-\alpha)\beta\delta Q\sigma}{\alpha R(M, N) + (1-\alpha)\beta\delta}\right), \quad (25)$$

with

$$R(M, N) = \frac{N}{N+1} + \frac{0.5N}{((1+N)^2 M)}, \quad (26)$$

(recall that $r = 1$, which implies $C = 1$). Using the previous equation, we obtained that:

$$\lim_{N \rightarrow 0} \Pi_s(\bar{a}, \gamma^*(\bar{a}), N) = \alpha Q\sigma,$$

and

$$\lim_{N \rightarrow +\infty} \Pi_s(\lambda^*(\hat{a}), \gamma^*(\hat{a}), \hat{a}, N) = 0.$$

Since Π_s decreases with N , it follows that for all $K \in (0, \alpha Q\sigma)$, there exists a unique N^* that solves:

$$\Pi_s(\lambda^*(\hat{a}), \gamma^*(\hat{a}), \hat{a}, N^*) = K. \quad (27)$$

N^* is the equilibrium number of speculators (again ignoring the integer problem). In what follows we study the effects of a change in some of the exogenous variables on both the equilibrium number of speculators and the equilibrium spread. The results will be used to formulate the empirical predictions that are derived in the next section.

Corollary 4 : *In equilibrium,*

- *When the average order size submitted by liquidity traders increases (δ increases), the number of speculators increases whereas the spread posted by the dealers decreases.*

- *When the number of dealers increases, the number of speculators increases whereas the spread posted by the dealers decreases.*
- *When β increases, the number of speculators increases whereas the spread posted by the dealers decreases.*

The economic reasoning behind these 3 statements is the same. Let us consider, for instance, an increase in the average order size of a liquidity trader, i.e., δ . The first effect of such an increase is that the spread posted by the dealers decreases, other things equal. But for this reason, more speculators are active (See Equation (24)). This counterbalances the initial effect of the change in the average order size on the spread but never sufficiently to ultimately result in an increase of the spread. We can use exactly the same argument in the case of an increase in M or an increase in β .

Corollary 5 : *In equilibrium:*

- *When the maximum order size for which dealers' quotes are binding increases, the number of speculators increases and the spread increases.*
- *When σ increases, the number of speculators increases and the spread increases.*

When Q increases, the expected gain for a speculator increases, other things equal (see Equation (24)). Thus the number of speculators increases. This effect however increases the cost of market making. The increase in the number of speculators increases the probability of a dealer being picked off. It follows that the equilibrium spread becomes larger. This counterbalances the initial effect of the increase in Q on the number of active speculators. But this feedback effect is never sufficiently large to result in a decrease in this number.¹³ The intuition for the effect of a change in σ is similar.

Dealers' quoted spreads and the number of speculators are jointly determined by the same set of variables. They do not necessarily evolve in the same direction. In fact, the two previous corollaries show that the quoted spread and the number of speculators can be negatively or positively related. For empirical analyses, our results suggest that in

¹³Note that we take δ as given in the analysis. This means that we assume that when Q increases, the average size of the order submitted by the liquidity trader increases as well by a factor δ . If the average size of the order submitted by liquidity traders is fixed, the effect described in the proposition is reinforced since an increase in Q as a direct, positive, effect on the spread.

regressions of a proxy for the level of SOES bandit activity, one should not include the quoted spread in the set of independent variables.

The unconditional probability of observing a trade initiated by a SOES bandit is given by $\alpha \text{Prob}(f \in \mathcal{N}) = \frac{\alpha N^*}{N^* + 1}$. Clearly this probability increases with the number of speculators. Thus it is determined by the same variables as the number of speculators and it behaves as the number of speculators when these variables change. This remark is used in the next section.

5 Empirical Implications

In this section, we briefly describe how our results relate to the existing empirical literature.

Harris and Schultz (1997) focus on the effects of the reduction in the maximum SOES order size from 1000 to 500 shares that occurred in 1994. Their proxy for SOES bandits' trades is a cluster of maximum-sized trades that occur at the same ask price. They find that the cross-sectional mean number of clusters per day declines significantly after the reduction in the maximum SOES order size. This finding is consistent with the first part of Corollary 5.¹⁴ The authors find that even if the change in the maximum SOES order size led to a dramatic decrease in clustered trades and therefore presumable to a decrease in market maker losses there was no significant change in the spread. This contradicts the prediction of our model, see Corollary 5.¹⁵

In a cross-sectional study, Kandel and Marx (1998) analyze the determinants of a proxy for SOES bandit activity¹⁶. They do not find that volatility is a significant determinant of this activity. The number of market makers has a positive impact on SOES bandit activity, as predicted by our model, but it is significant in only 2 of the 6 regressions they run. The authors also include the number of quote updates per day, which may be strongly

¹⁴The 'length' of these 'clusters' is related to the how quickly the SOES bandits and the market makers react to a given trading opportunity, say, a public announcement. If the market maker are relatively slow to act we expect to observe substantially positive autocorrelation at long lags for the maximum-size orders. This relationship could be used directly to identify some of the parameters describing the monitoring activity of speculators and market makers in our model. Similarly the quote update made by market makers without any preceding trades are related to monitoring activity and could be used to identify these parameters.

¹⁵Benston and Wood (1998) results suggest that the spread is positively related to SOES trading activity.

¹⁶Kandel and Marx (1998) use data for the months of January and February 1994. The reduction in the maximum trade size in SOES occurred on January 31, 1994. For January, their proxy is the difference between the average daily proportion of trades of 1000 shares in January and this proportion in February. For February, their proxy is the difference between average daily proportion of trades of 500 shares trades in February and this proportion in January.

correlated with the number of market makers.

In a time-series study, Battalio *et al.* (1997) find that high levels of volatility precede trades by SOES bandits. They do not find any significant effect for the number of market makers. It is worth stressing that no study control for the average size of a SOES trade relative to the maximum-sized order. In order to test the predictions of our model directly it is necessary to be able to quantify the extent of market monitoring. The autocorrelation in maximum-size SOES trades as well as the frequency and correlation between market makers' quote updates, without preceding trades, could be used to characterize the monitoring in the market. Combined with the cross-sectional data on bid-ask spreads this may permit identification of the implied costs of monitoring. We plan to explore these possibilities in future versions of the paper.

6 Conclusions

In this paper, we present a simple model of price formation in a market where monitoring of news is costly. Our model is designed to capture key features of Nasdaq's Small Order Execution System (SOES). In our model, market makers post firm quotes and select how intensively they monitor the arrival of news. Due to monitoring costs market makers optimally do not monitor the news arrival process continuously. This implies that in the event of news it is possible that all market makers' quotes will be 'stale'. The possibility of 'stale' quotes create profit opportunities for a second group of agents referred to as speculators, also monitor the news arrival but do not post quotes. The speculators exploit the profit opportunities created by imperfect monitoring by trading with the market makers. The market makers losses from trading with the speculators are offset by gains from trading with liquidity traders. The monitoring of news is a *public good* for the market maker and they therefore monitor less than they would in the absence of this externality. The speculators do not face the same problem since they individually profit from their own monitoring efforts. Thus, both the monitoring costs per se and the public good problem provide rationales for profitable trading by speculators in equilibrium. Differences in monitoring skills or costs are not necessary to induce speculators to enter the market. Our results provide a rational for the existence of SOES bandits on Nasdaq.

The SOES has generated intense regulatory interest concerning the alleged effects of the SOES on trading costs, volatility, liquidity, entry/exit of market makers, and the quality

of the price discovery process.¹⁷ Our model contributes to this debate by formally showing how speculators can improve the price discovery process in a market with costly monitoring due to externalities in monitoring. We also show that the presence of speculators creates an adverse selection problem for the market makers, which increase the trading costs. Thus, based on our theoretical model the trade-off between the spread and the quality of price discovery is one relevant issue for the SOES debate.

We believe that our model can be extended to address several other interesting questions regarding the impact of automated execution of trades on trading costs, liquidity, and the price discovery process. One effect of the emergence of the SOES bandits is that market makers have been forced to make additional investment in software to monitor their quotes and limit their exposure to SOES bandits. Increased automation of the quote setting process may decrease the market makers' exposure to the SOES bandits but it is also likely to alter the market makers' incentives to monitor the arrival of new information. It is also important to notice that we have focused on public information alone in this model. The incentives to gather costly information that is not public may also be altered if SOES bandits, as it sometimes loosely claimed, are able to 'front run' other traders.

While we have concentrated on the key features of the SOES in our model it is clear that several other markets also allow automated execution of orders as well as increasingly easy access to quotes for public investors. For example, limit order markets¹⁸ allow all participants to execute orders immediately against the limit order book. Some markets, e.g., Stockholm Stock Exchange, also allow individual investors direct access to the limit order book through a brokerage firm. Given the differences between Nasdaq and most limit order markets, for example, time priority rules and a centralized order book, the impact of costly monitoring on trading costs and price discovery in these markets is an interesting question for future research.

¹⁷See, for example, the GAO report on 'The Effects of SOES on the Nasdaq Market' for a summary of the main arguments and important SOES-related events.

¹⁸For example, Paris Bourse, Toronto Stock Exchange, and Stockholm Stock Exchange.

7 Appendix

Proof of Lemma 1.

Dividing Equation (11) by Equation (12), we find that $\lambda^*(a^*)$ and $\gamma^*(a^*)$ must satisfy:

$$\frac{N\gamma^*}{M\lambda^* + (N-1)\gamma^*} = (rM)\frac{\lambda^*}{\gamma^*} \quad (28)$$

It follows that it must be the case that $\lambda^* = \frac{C}{M}\gamma^*$ where C is given by:

$$C = 0.5\left(\sqrt{(N-1)^2 + \frac{4N}{r}} - (N-1)\right)$$

Replacing λ^* by $\frac{C}{M}\gamma^*$ in Equation (12), we find that γ^* solves:

$$\frac{\alpha(C + (N-1))Q(\sigma - a^*)}{\gamma^*(C + N)^2} = c_s\gamma^*$$

Solving for γ^* , we get the closed form solution for γ^* as announced in the Lemma.
Q.E.D.

Proof of Corollary 1.

Recall that:

$$R(r, M, N) = \frac{N}{C + N} + \frac{1}{2} \frac{CN}{M(C + N)^2} \quad (29)$$

Thus we obtain:

$$\frac{\partial R}{\partial r} = -\frac{\partial C}{\partial r}(C + N)^{(-2)} \left(N - \frac{N}{2M} + \frac{NC}{M(C + N)} \right)$$

Since $\frac{\partial C}{\partial r} < 0$, we obtain that $\frac{\partial R}{\partial r} > 0$. Since $\hat{a}$ increases with R , it follows that:

$$\frac{\partial \hat{a}}{\partial r} > 0$$

C becomes infinite when r goes to zero and goes to zero when r becomes infinite. Consequently $\lim_{r \rightarrow 0} R(M, N, r) = 0$ and $\lim_{r \rightarrow +\infty} R(M, N, r) = 1$. It follows that:

$$\lim_{r \rightarrow 0} \hat{a}(r, M, N, \delta) = 0$$

and

$$\lim_{r \rightarrow +\infty} \bar{a}(r, M, N, \delta) = \frac{\alpha\sigma}{\alpha + (1 - \alpha)\beta\delta}$$

The same type of computations yield the results for $\hat{a}$. *Q.E.D*

Proof of Corollary 2.

Since $C = 1$ when $r = 1$, we obtain:

$$R(1, M, N) = \frac{N}{N+1} + \frac{N}{2M(N+1)^2}$$

$$\bar{R}(1, M, N) = \frac{N}{N+1} + \frac{N(M+1)}{2M(N+1)^2}$$

It is straightforward that $R(1, M, N)$ and $\bar{R}(1, M, N)$ increase with N and decrease with M . Using the fact that $\hat{a}$ and $\bar{a}$ increase with R and $\bar{R}$, respectively, we obtain the corollary. *Q.E.D*

Proof of Corollary 3.

Take the first derivatives of $\hat{a}$ and $\bar{a}$ with respect to q , β and α . The result is then immediate.

Proof of Corollary 4.

Using Equation (25), it is straightforward that:

$$\frac{\partial \Pi_s}{\partial \delta} > 0 \tag{30}$$

Since Π_s decreases with N , this implies that:

$$\frac{\partial N^*}{\partial \delta} > 0 \tag{31}$$

Now suppose that $\frac{\partial \hat{a}(1, M, N^*(\delta), q)}{\partial \delta} < 0$. Using Equation (24), this would imply that $\frac{\partial \Pi_s}{\partial \delta} < 0$. But this contradicts Equation (30). It follows that:

$$\frac{\partial \hat{a}(1, M, N^*(\delta), \delta)}{\partial \delta} > 0 \tag{32}$$

Q.E.D.

Proof of Corollary 5.

Using Equation (25), it is straightforward that:

$$\frac{\partial \Pi_s}{\partial Q} > 0 \tag{33}$$

It follows that N^* increases with Q . Since $\hat{a}$ increases with the number of speculators, it follows that $\hat{a}$ increases with Q . In the same way, we obtain that $\frac{\partial \Pi_s}{\partial \sigma} > 0$. Thus N^* increases with σ . Furthermore, for a given number of speculators, the spread posted by the dealers increases with σ . Consequently $\hat{a}$ increases with σ . *Q.E.D.*

References

- [1] Barclay, M., Christie, W. , Harris J. , Kandel, E., and Schultz, P., 1998, The Effect of Market Reform on the Trading Costs and Depths of Nasdaq Stocks, forthcoming, *Journal of Finance*.
- [2] Battalio, R., Hatch, B., and Jennings, R., 1997, SOES Trading and Market Volatility, *Journal of Financial and Quantitative Analysis*, 32, 225-238.
- [3] Benston, G., and Wood, R., 1998, Day Trading on Nasdaq's Automatic Small Order Execution System (SOES): Adverse Selection, working paper, University of Memphis.
- [4] Christie, W., and Schultz, P., 1994, Why Do NASDAQ Market Makers Avoid Odd-eight Quotes?, *Journal of Finance*, 49, 1813-1840.
- [5] Copeland, T., and Galai, D., 1983, Information Effects on the Bid-Ask Spread, *Journal of Finance*, 38, 1457-1469.
- [6] General Accounting Office, 1998, The Effects of SOES on the Nasdaq Market, United States General Accounting Office Report 98-194.
- [7] Harris, J., and Schultz, P., 1997, The Importance of Firm Quotes and Rapid Executions: Evidence from the January 1994 SOES rules change, *Journal of Financial Economics*, 45, 135-166.
- [8] Harris, J., and Schultz, P., 1998, The Trading Profits of SOES Bandits, *Journal of Financial Economics*, 50, 39-62.
- [9] Hinden, S., 1994, Nasdaq's Big Guns Send Trading Bandits Packing, Washington Post, February 7, 1994.
- [10] Houtkin, H., 1998, Secrets of the SOES Bandit, McGraw-Hill.
- [11] Kandel, E., and Marx, L., 1998, Odd-eight Avoidance as a Defense Against SOES Bandits, forthcoming, *Journal of Financial Economics*.
- [12] Kleidon, A., and Willig, R., 1995, Why Do Christie and Schultz Infer Collusion from Their Data?, working paper, Cornerstone Research, Menlo Park, CA.

- [13] Whitcomb, D., 1998, The NASDAQ Small Order Execution System: Myth and Reality, testimony before the House Committee on Commerce, Subcommittee on Finance, August 3, 1998.

Table 1: Payoffs

B is the set of dealers posting the best quote in the market and a^* denotes the best quote. The number of dealers posting the best quote is m .

Payoffs of dealer i

Event	$\{V_1=0, i \in B\}$	$\{V_1=0, i \notin B\}$	$\{V_1=\sigma, f \in M\}$	$\{V_1=\sigma, f \in N\}$
Payoff	$(\delta qa^*)/m - C(\lambda_i)$	$-C(\lambda_i)$	$-C(\lambda_i)$	$i \in B$
				$-C(\lambda_i)$

Payoff for Speculator j.

Event	$\{V_1=0\}$	$\{V_1=\sigma, f \in M\}$	$\{V_1=\sigma, f =j\}$
Payoff	$-C(\gamma_j)$	$-C(\gamma_j)$	$Q(\sigma - a^*) - C(\gamma_j)$

Figure 1: Timing of the trading game

