Market Making with Costly Monitoring: An Analysis of the SOES Controversy*

Thierry Foucault
HEC, School of Management and CEPR

Ailsa Röell
Princeton University and CEPR

Patrik Sandås
University of Pennsylvania and CEPR

January 13, 2002

*An earlier draft of this article was entitled "Imperfect Market Monitoring and SOES Trading." We have benefited from comments by Dan Bernhardt, Bruno Biais, Hans Degryse, Frank De Jong, Richard Evans, Terrence Hendershott, Burton Hollifield, Eugene Kandel, Ken Kavajecz, Gaëlle Lefol, Maureen O'Hara (the editor), Barbara Rindi, Andrei Simonov, Chester Spatt, Matthew Spiegel, Erik Theissen, Ingrid Werner, two anonymous referees, and seminar participants at University of California at Berkeley, Bocconi University, Carnegie Mellon University, Frankfurt University, HEC School of Management, Leuven University, London School of Economics, Wharton, the Security and Exchange Commission, the meetings of the Western Finance Association (1999), the European Finance Association (2000), the European Summer Symposium in Financial Markets (1999), the JFI Symposium in Boston (2000), and the E-finance conference organized by The Federal Reserve Bank of Atlanta (2000). Financial support from the Nasdaq-Amex Center for Financial Research and the Foundation HEC is gratefully acknowledged. All errors are ours. Address all correspondence to: Patrik Sandås, Finance Department, The Wharton School, University of Pennsylvania, Steinberg Hall – Dietrich Hall, Philadelphia, PA 19104. E-mail: sandas@wharton.upenn.edu

Abstract

We model how intensively dealers monitor public information to avoid being picked off by professional day traders when monitoring is costly. Price competition among dealers is hampered by their incentive to share monitoring costs. The risk of being picked off by the day traders makes dealers more competitive. The interaction between these effects determines whether a firm quote rule improves trading costs and price discovery. Our empirical results support the prediction that professional day traders prefer stocks with small spreads, but offers less support for the prediction that their trading leads to wider dealer spreads.

Keywords: Market Making, Monitoring, Bid-Ask Spread, SOES, Nasdaq.

Introduction

Nasdaq's Small Order Execution System (SOES) allows brokerage firms to execute small orders automatically at the best quotes posted by Nasdaq dealers. Participation in SOES and in its new incarnation SuperSoes is mandatory for all dealers, who must post firm quotes for a minimum quantity, fixed by Nasdaq. Although it was intended for retail investors, SOES mainly attracted professional day traders (labeled SOES "bandits" by Nasdaq dealers). The bandits make money by detecting short-term price trends and trading before all dealers have incorporated this new information into their quotes.[1] SOES bandits and their alleged adverse impact on Nasdaq trading costs, liquidity, and volatility has been the subject of a long and heated policy debate.[2]

Harris and Schultz (1998) examine SOES bandit trading strategies and profits using data from two brokerage firms that cater to bandits. They conclude:

> *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. In response, Nasdaq market makers have expended considerable effort to eliminate SOES bandits through regulation. They have invested hundreds of thousands of dollars in proprietary software to update quotes when bandits trade against them. Why do not market makers just hire traders to keep track of other dealers' quotes, Instinet quotes and SelectNet quotes and update their own prices in a more timely fashion?* [p. 61]

How do market makers combine software and labor to ensure that their quotes reflect all available information? This question is central to understanding the profitability of SOES bandits. How should traders at market making firms, or bandits for that matter, allocate their time between processing relevant information for a given stock and the most valuable alternative use of their time? In other words, how much should they invest in costly monitoring. We attempt to address these questions by developing a model of market making where both market makers and bandits must choose how intensively to monitor information. We consider two forms of monitoring: (i) *news monitoring* and (ii) *quote monitoring*. News monitoring entails monitoring, for example, public announcements, whereas quote monitoring is limited to monitoring other dealers' quote updates. News monitoring is costly because the correct interpretation of, say, a corporate announcement requires human attention. In contrast, the use of software makes the marginal cost of quote monitoring very low.

In our model, dealers post firm quotes and select how intensively they monitor news. They never monitor news continuously because it is costly to do so. At times dealer quotes do not reflect all public information because monitoring is imperfect. These so called *stale quotes* provide profit opportunities for the bandits. Bandits also monitor news and quote updates with a view to detect these opportunities and exploit them by trading with dealers before they update their quotes. In equilibrium, bandits' expected trading profits are positive. Dealers offset their losses to the bandits by gains from trading with liquidity traders.

We obtain three main results. First, news monitoring by one dealer can generate either a *positive* or a *negative externality* for the other dealers. By monitoring quote updates, a dealer can free ride on the efforts that his competitors exert to monitor the flow of information. Thus, monitoring gives rise to a positive externality. On the other hand, bandits may discover that some dealers' quotes are stale by observing other dealers' updating their quotes. This introduces a negative externality of monitoring. Whether the positive or the negative externality is stronger depends on how quickly the dealers react to quote updates. Second, these externalities influence the dealers' bidding behavior. The positive externality induces dealers to match the best quotes rather than to undercut them. This effect produces multiple equilibria in which dealers earn strictly positive expected profits. In contrast, the negative externality generates an equilibrium with very low liquidity in which only one dealer posts the inside spread and makes zero expected profits. Third, the bandits' ability to profit from the information in quote updates hinges on the fact that quotes are firm and order execution is automatic. We show that relaxing the firm quote rule, e.g., allowing the dealers an option to "back away" from their quotes, can increase (decrease) spreads and slow down (speed up) price discovery, depending upon which equilibrium is obtained.

Despite the frequent claims that bandits have an adverse impact on trading costs and liquidity, there is surprisingly little empirical evidence to support this claim. It is difficult to obtain direct evidence on this effect because the spread and the level of bandit activity are interdependent. An increase in the spread triggers the exit of some bandits, whereas an increase in the number of bandits triggers a widening of the spread. We disentangle this interdependence by formulating a two-equation model, which we use to test whether more bandit activity leads to wider spreads. Consistent with the predictions of our theoretical model, we find, for two different samples, that a wider spread is associated with less SOES bandit activity. However, only for a sample of the most actively traded stocks do we find that a higher level of SOES bandit activity is associated with a wider spread, and even in this case the effect is statistically significant only at the 10% level. For a second sample of less actively traded stocks we cannot reject the null hypothesis that bandit activity has no effect on the bid-ask spread.

Battalio, Hatch, and Jennings (1997) show that SOES bandits speed up the price discovery process and are more likely to trade in volatile periods. We obtain theoretical and empirical results consistent with their findings. Harris and Schultz (1997) report evidence consistent with a reduction in SOES bandit activity following a reduction in the minimum depth from 1000 to 500 shares. In our model, a decrease in the mandatory quoted depth causes fewer bandits to enter and thus tightens the spread. Our empirical results provide strong support for the former prediction but only weak support for the latter one. We also show theoretically that another effect of a reduction in the minimum quoted depth is to slow down price discovery.

Our model is related to Copeland and Galai (1983), who analyze the free-trading option aspect of fixed quotes. We show how the free-trading option problem arises in equilibrium as a result of costly monitoring. Kandel and Marx (1999) develop a theoretical model to study whether odd-eighth avoidance is a rational response by Nasdaq dealers to SOES bandits. In their model the profit opportunities of the SOES bandits are *implicitly* assumed to be due to imperfect monitoring by the dealers. We *explicitly* model how stale quotes or profit opportunities may arise. Kumar and Seppi (1994) model how index arbitrageurs learn information from quote updates. In their model the index arbitrageurs *always* observe quote updates more quickly than do dealers, which is not the case in our analysis for the bandits vis-à-vis the dealers.

The article is organized as follows. In Section 1, we develop the model. In Section 2, we present the optimal monitoring strategies and the information externalities. In Section 3, we derive the equilibrium spreads given the monitoring strategies. In Section 4, we examine how relaxing the firm quote rule affects market quality. In Section 5, we derive empirical implications for the level of SOES bandit activity and the bid ask spread. In Section 6, we estimate a two-equation model of SOES bandit activity and the bid ask spread. In Section 7, we summarize our conclusions. All proofs are in the Appendix.

1 The Model

1.1 The Structure of the Trading Game

There is a single risky asset with a liquidation value, $\tilde{V}$. At the beginning of the trading round, the expected liquidation value is v_0. There are three types of traders: (i) $M \geq 2$ *dealers*, (ii) $N \geq 1$ *bandits*, and (iii) *liquidity traders*. All traders are risk neutral.

A trading round consists of three stages, as illustrated in Figure 1. In the *quoting stage*, dealers simultaneously quote their spreads, $\{S_i\}_{i=1}^{i=M}$. Dealer i's bid quote is $b_i = v_0 - \frac{S_i}{2}$ and his ask quote is $a_i = v_0 + \frac{S_i}{2}$. We denote the *inside spread* (the smallest posted spread) by S_b. The number of dealers posting the inside spread is denoted M_b. Dealer quotes are firm for up to Q shares, the minimum quoted depth. In the *monitoring stage*, after observing the quotes, the dealers and the bandits choose their monitoring levels. This choice determines the probability that a trader is the first to discover an innovation in the asset value. In the *trading stage*, one of the following events occurs. With probability $\alpha < 1$, there is an innovation in the asset value. In this case the new asset value is either $v_1 = v_0 + \frac{\sigma}{2}$ or $v_1 = v_0 - \frac{\sigma}{2}$ with equal probabilities. Conditional on an innovation, a bandit may buy or sell the asset before dealers update their quotes. With probability $(1 - \alpha)$, there is no innovation. In this case, with probability $\beta > 0$, a buy or a sell order is submitted by a liquidity trader, with equal probabilities. The expected size of the liquidity trader's order is δQ. With probability $(1 - \beta)$, no order is submitted.

Market orders are evenly split among the dealers posting the *best quotes*. A dealer trades $\frac{\delta Q}{M_b}$ shares of a liquidity trader's order. A bandit places at most L orders of size Q and cannot place more orders than the total quoted depth, M_bQ. Hence the total size of a bandit trade is,

$$Q^s(M_b) = Min\{M_b, L\}Q = Min\{1, L/M_b\} \times M_bQ. \quad (1)$$

Each dealer trades $\frac{Q^s(M_b)}{M_b} = Min\{1, L/M_b\}Q$ shares of a bandit's order. For conciseness, we denote the portion of the quoted depth which is exposed to bandits by $x^s(M_b) = Min\{1, L/M_b\}$. We refer to $x^s(M_b)$ as the dealer's participation rate in bandit trades.[3]

1.2 News Monitoring and Quote Monitoring

Dealers and bandits become aware of new information by directly monitoring the information flow, an activity that we call *news monitoring*.[4] We model news monitoring as follows. Let $\lambda_i(\geq 0)$ be the *monitoring level* of dealer i and let $\gamma_j(\geq 0)$ be the monitoring level of bandit j. If new information arrives, the probability that a trader, say m, is first to observe news is denoted by $Prob(f = m)$. This probability depends on the monitoring levels as follows

$$Prob(f = i) \equiv P(\lambda_i) \equiv \frac{\lambda_i}{\lambda_i + \sum_{m \neq i} \lambda_m + \sum_j \gamma_j} \quad \forall i \in \mathcal{M}, \tag{2}$$

$$Prob(f = j) \equiv P(\gamma_j) \equiv \frac{\gamma_j}{\gamma_j + \sum_{k \neq j} \gamma_k + \sum_i \lambda_i} \quad \forall j \in \mathcal{N}, \tag{3}$$

where $\mathcal{M}$ denotes the set of dealers and $\mathcal{N}$ denotes the set of bandits. We assume $P(0) = 0$ and $P(+\infty) = 1$. A zero monitoring level corresponds to *no monitoring of news at all.* Conversely, an infinite monitoring level corresponds to *continuous* news monitoring. For any intermediate level news monitoring is imperfect. The probability that a trader is first to observe an innovation increases in his own monitoring level and decreases in the aggregate monitoring level. Monitoring requires effort and the monetary disutility associated with this effort is captured by a strictly increasing and *strictly convex* cost function $\Psi(l)$. We assume that

$$\Psi(l) = \frac{cl^2}{4}, \tag{4}$$

where l denotes the monitoring level and the parameter $c > 0$ determines the scale of the monitoring cost for a given monitoring level.[5] Bandits and dealers *simultaneously* choose their monitoring levels, after observing the inside spread. We denote the vector of the dealers' monitoring levels by $\lambda(S_b, M_b) = (\lambda_1(S_b, M_b), \ldots, \lambda_{M_b}(S_b, M_b))$. Dealers posting wider spreads than the inside spread choose not to monitor, since orders are only routed to the dealers at the inside. Analogously, $\gamma(S_b, M_b) = (\gamma_1(S_b, M_b), \ldots, \gamma_N(S_b, M_b))$ denotes the bandits' monitoring levels.

Dealers and bandits also monitor quote updates (*quote monitoring*). Dealers use the information revealed by quote changes to update their quotes. Bandits use quote updates to detect stale quotes.[6] We assume that when a dealer is first to update his quotes, there is a probability Φ that one bandit reacts to this quote update before the other dealers react. In this case, each bandit has an equal probability ($1/N$) of reacting first. With probability $(1 - \Phi)$, the other dealers update their quotes before any one of the bandits react. Thus, Φ measures the relative advantage of the bandits in quote monitoring (if $\Phi = 0$, dealers always react more quickly than bandits and vice versa if $\Phi = 1$). Quote monitoring is pointless when there is only one dealer at the inside. Hence, for $M_b = 1$, we set $\Phi = 0$.

In practice, bandits and dealers use software that alerts them to quote updates in different securities. For this reason, we assume that Φ does not depend on the levels of news monitoring. One likely determinant of this probability, which is not examined here, is the fixed cost of the trading technology used. Other determinants include rules concerning firm quotes and automatic quote updates. We return to the firm quote rule in Section 4.

The optimal response for the dealers and the bandits in the trading stage is as follows. If a dealer is first to observe the new information, he revises his quotes. If his competitors react to this quote update before the bandits, they revise their quotes as well. If a bandit is first to react to a quote update or to observe new information, she submits buy (sell) orders when she observes a good (bad) signal.[7] Tables 1 and 2 list the payoffs for the dealers and the bandits, for a given spread and fixed monitoring levels.

1.3 Discussion of the Assumptions

The quantity, Q, corresponds to the minimum quoted depth in the SOES system. Nasdaq dealers execute orders at their posted quotes that are larger than the minimum quoted depth. SOES bandits typically do not take part in these trades since they are negotiated by phone. This slows down the execution process and dealers can back away from their quotes upon realizing that a bandit is trying to initiate a trade (see Harris and Schultz (1997) and Houtkin (1998)). Accordingly, the size of liquidity trades can be larger than Q (i.e., $\delta > 1$). NASD rules prohibit individual bandits from initiating more than one position (i.e., $L = 1$) in the same stock within a five minute interval. By varying L we can study the effects of relaxing this rule. It is worth stressing that variations in L are not equivalent to variations in δ. The reason is that the size of bandits' trades depend on the total quoted depth but not the size of liquidity trades. Hence a decrease in the number of dealers at the inside necessarily enlarges a dealer's participation rate in liquidity trades but may leave unchanged his participation rate in bandit trades (if L is large enough).

In some equilibria **only one** dealer can profitably post the inside spread. In these equilibria sidelined dealers are exposed only to bandits, since liquidity traders are executed at the inside quotes. Hence the sidelined dealers widen their spreads to avoid being picked off. In order to account for this reaction within our *static* model, we simply assume that orders are only routed to the dealers posting the inside spread.[8] This is in fact the case in SOES.

We assume that bandits unwind their positions at the mid-quote (v_1) subsequent to information arrival. Bandits frequently unload their positions on Selectnet or Instinet and trade within the quoted bid-ask spread. In fact Harris and Schultz (1998) find that when bandits lay off their positions, they trade at the spread mid-point or at a more favorable price in 90% of the cases. More generally, we could assume that bandits pay a fixed fraction τ of the spread when they close out their positions (as in Kandel and Marx (1999)). They would then gain $(\sigma-(1+\tau)S)/2$ instead of $(\sigma - S)/2$ when they initiate a trade. This just scales up the effect of the spread on bandits' payoffs and would not qualitatively affect our results.

Finally, the probability of a liquidity trade after an informational event is assumed to be zero. This assumption could easily be relaxed. Increasing the probability of a liquidity trade after an innovation reduces the risk of being picked off for the dealers and is tantamount to a decrease in the probability of an informational event (α).

2 Monitoring

We focus on perfect equilibria of the trading game. In a perfect equilibrium, **(i)** traders' monitoring strategies ($\lambda^*(S_b, M_b)$ and $\gamma^*(S_b, M_b)$) form a Nash equilibrium given the outcome of the quoting stage, and **(ii)** dealers' quotes form a Nash equilibrium, given the monitoring strategies. We start by analyzing the monitoring strategies.

2.1 Monitoring Externalities

In this section, we show that news monitoring by one dealer can generate a positive *or* a negative externality for the other dealers. Consider one dealer, say i. There are two ways dealer i can be picked off. In the first case, a bandit reacts first to news. Using Equation (3), this event occurs with probability

$$Prob(f \in \mathcal{N}) = \frac{\gamma_A}{\lambda_A + \gamma_A}, \tag{5}$$

where $\lambda_A \equiv \sum_i \lambda_i$ and $\gamma_A \equiv \sum_j \gamma_j$ are the aggregate monitoring levels. In the second case, a different dealer (i.e., not dealer i) observes the news and updates his quotes, and a bandit is first to react to the quote update. The probability of this event is $\Phi Prob(f \in \mathcal{M}_b \backslash i)$. Using Equation (2), we obtain

$$Prob(f \in \mathcal{M}_b \backslash i) = \frac{\sum_{m \neq i} \lambda_m}{\lambda_A + \gamma_A}. \tag{6}$$

Let $\Pi_d(\lambda_i, \lambda_{-i}, \gamma)$ be dealer i's expected profit for given levels of monitoring, λ_{-i} and γ, for the other dealers and the bandits, respectively. Using the payoffs listed in Table 1, we get the following expression for dealer i's expected profit:

$$\begin{aligned} \Pi_d(\lambda_i, \lambda_{-i}, \gamma) \quad = \quad & -\alpha \left[x^s(M_b) Prob(f \in \mathcal{N}) + x^s(M_b - 1) \Phi Prob(f \in \mathcal{M}_b \backslash i) \right] \frac{(\sigma - S_b)Q}{2} \\ & + \left[(1-\alpha)\beta \right] \frac{S_b \delta Q}{2M_b} - \Psi(\lambda_i) \quad \forall M_b \geq 2. \end{aligned} \tag{7}$$

The first term represents dealer i's expected loss when he is picked off. The second term corresponds to dealer i's expected gain from trading with a liquidity trader. The last term is the monitoring cost incurred by dealer i. The expected loss for dealer i is affected by the monitoring levels chosen by himself as well as the levels chosen by the other dealers.

Proposition 1. *Consider two dealers i and m who are posting the inside spread. There exists a cut-off $\bar{\Phi} = \frac{\gamma_A x^s(M_b)}{(\gamma_A + \lambda_i) x^s(M_b - 1)} < 1$ such that news monitoring by dealer m is a:*

1. *Positive externality for dealer i, or $\frac{\partial \Pi_d(\lambda_i, \lambda_{-i}, \gamma)}{\partial \lambda_m} \geq 0$, if $\Phi \leq \bar{\Phi}$.*

2. *Negative externality for dealer i, or, $\frac{\partial \Pi_d(\lambda_i, \lambda_{-i}, \gamma)}{\partial \lambda_m} < 0$, if $\Phi > \bar{\Phi}$.*

An increase in news monitoring by dealer m increases the probability that dealer m will be first to observe news. This indirectly benefits dealer i, since a quote update by dealer m *signals* to dealer i that his quotes are stale. Thus, an increase in news monitoring by dealer m reduces the risk of dealer i being picked off by bandits (that is $\frac{\partial Prob(f \in \mathcal{N})}{\partial \lambda_m} < 0$). This is the source of the positive externality. There is, however, a second effect, since bandits also monitor quote updates. An increase in news monitoring by dealer m increases the risk of dealer i being picked off by bandits who discover stale quotes through quote monitoring (that is $\frac{\partial Prob(f \in \mathcal{M}_b \backslash i)}{\partial \lambda_m} > 0$). This is the source of the negative externality. If dealer i reacts sufficiently quickly to dealer m's quote updates ($\Phi \leq \bar{\Phi}$), the reduction in the picking off risk due to news monitoring is larger than the increase in the picking off risk due to quote monitoring. If bandits are relatively quicker ($\Phi > \bar{\Phi}$), the reverse is true.

2.2 Equilibrium in the Monitoring Stage

Dealer i chooses the monitoring level that maximizes his expected profit. Using Equation (7), the first order condition is

$$-\alpha \left[x^s(M_b) \frac{\partial Prob(f \in \mathcal{N})}{\partial \lambda_i} + x^s(M_b - 1)\Phi \frac{\partial Prob(f \in \mathcal{M}_b \backslash i)}{\partial \lambda_i} \right] \frac{(\sigma - S_b)Q}{2} = \Psi^{'}(\lambda_i).$$

The terms inside the brackets measure the marginal reduction in the probability of being picked off due to increased monitoring by dealer i. Using Equations (5) and (6), we rewrite this as[9]

$$\frac{\alpha(\sigma - S_b)Q}{2(\lambda_A + \gamma_A)^2} \left[x^s(M_b)\gamma_A + x^s(M_b - 1)\Phi \sum_{m \neq i} \lambda_m \right] = \Psi^{'}(\lambda_i). \tag{8}$$

Using the payoffs listed in Table 2, we obtain the following expression for the expected profit of bandit j, $\Pi_s(\gamma_j, \lambda, \gamma_{-j})$,

$$\Pi_s(\gamma_j, \lambda, \gamma_{-j}) = \frac{\alpha(\sigma - S_b)}{2} \left[Prob(f = j)Q^s(M_b) + \frac{\Phi Prob(f \in \mathcal{M}_b)}{N} Q^s(M_b - 1) \right] - \Psi(\gamma_j), \tag{9}$$

where $Prob(f \in \mathcal{M}_b) = \frac{\lambda_A}{\lambda_A + \gamma_A}$ is the probability that a dealer is first to observe new information. The term inside brackets is the expected trade size for a bandit. Bandits exploit stale quotes either by (i) learning about news first, or (ii) reacting quickly to quote changes. In the first case she trades $Q^s(M_b)$ shares whereas in the second case she trades $Q^s(M_b - 1)$ shares. Bandit j chooses the monitoring level that maximizes her expected profit. The first order condition is

$$\frac{\alpha Q^s(M_b)(\sigma - S_b)}{2(\lambda_A + \gamma_A)^2}\left[(\frac{N - h(\Phi, L)}{N})\lambda_A + \sum_{s \neq j}\gamma_s\right] = \Psi^{'}(\gamma_j), \tag{10}$$

where $h(\Phi, L) = \Phi\frac{Q^s(M_b-1)}{Q^s(M_b)} < 1$. A Nash equilibrium of the monitoring stage is a set of monitoring levels that solve Equations (8) and (10). This equilibrium is *symmetric* if all the traders of a given type choose the same monitoring level.

Lemma 1. *If there exists a Nash equilibrium in the monitoring stage, it is symmetric.*

Let λ^* (γ^*) be the monitoring level chosen by each dealer (bandit) in equilibrium. From Equation (4), we get that $\Psi^{'}(l) = cl/2$. Using this expression, we rewrite the system of Equations (8) and (10) characterizing traders' best responses as

$$\frac{\alpha(\sigma - S_b)Q}{(M_b\lambda^* + N\gamma^*)^2}\left[x^s(M_b)N\gamma^* + x^s(M_b - 1)\Phi(M_b - 1)\lambda^*\right] = c\lambda^*, \tag{11}$$

and

$$\frac{\alpha Q^s(M_b)(\sigma - S_b)}{(M_b\lambda^* + N\gamma^*)^2}\left[(\frac{N - h(\Phi, L)}{N})M_b\lambda^* + (N-1)\gamma^*\right] = c\gamma^*. \tag{12}$$

Solving this system of equations yields the equilibrium monitoring levels.

Proposition 2. *When M_b dealers post an inside spread $S_b \leq \sigma$, the equilibrium of the monitoring stage is unique and is characterized by the following monitoring levels for the bandits and the dealers:*

$$\gamma^*(S_b, M_b) = \left(\frac{N - h(\Phi, L)}{N + 1 - h(\Phi, L)}\right)\sqrt{\frac{\alpha Q^s(M_b)(\sigma - S_b)}{cN}}, \tag{13}$$

$$\lambda^*(S_b, M_b) = \sqrt{\frac{\alpha N Q^s(M_b)(\sigma - S_b)}{cM_b^2(N + 1 - h(\Phi, L))^2}} = \frac{N\gamma^*}{M_b(N - h(\Phi, L))}. \tag{14}$$

For these monitoring levels, the expected profits of the dealers and the bandits are

$$\Pi_d^*(S_b, M_b) = \frac{Q}{2}\left[-\alpha x^s(M_b)C(M_b, \Phi, L)(\sigma - S_b) + \frac{(1-\alpha)\beta\delta S_b}{M_b}\right], \tag{15}$$

$$\textit{with}\quad C(M_b, \Phi, L) \equiv \frac{N}{N + 1 - h(\Phi, L)} + \frac{N}{2M_b(N + 1 - h(\Phi, L))^2}. \tag{16}$$

$$\Pi_s^*(S_b, N) = \alpha Q^s(M_b)(\sigma - S_b)\left[\frac{2N(N+1-h(\Phi,L)) - (N-h(\Phi,L))^2}{4N(N+1-h(\Phi,L))^2}\right]. \tag{17}$$

The proposition reveals several interesting properties of the monitoring strategies. First, bandits and dealers always put some effort into news monitoring, i.e., $\gamma^* > 0$ and $\lambda^* > 0$. In particular, it is never optimal for bandits to entirely base their trading strategies on dealers' quote updates.[10] Second, the optimal monitoring levels decrease in the spread. When the dealers increase their spread, bandits monitor the market less intensively, since the potential profit from picking off dealers is lower. The dealers react by monitoring less.[11] The negative term in a dealer's expected profit is the expected trading loss to bandits ('the cost of market-making'). Part of this cost, $C(M, \Phi, L)$, reflects the joint effect of all traders' monitoring decisions on the probability of a dealer being picked off and his monitoring cost.

Lemma 2. *The component of the cost of market making which is determined by traders' monitoring decisions, $C(M, \Phi, L)$, increases with Φ.*

An increase in the bandits' relative advantage in quote monitoring implies a greater picking off risk for the dealers. They react by choosing higher monitoring levels, other things equal. But, in equilibrium, this is insufficient to fully counter-balance the increase in the risk of being picked off. Hence an increase in Φ results in a higher monitoring cost and a greater risk of being picked off. Lemma 2 follows. Proposition 2 also holds when only one dealer posts the inside spread ($M_b = 1$). In this case $\Phi = 0$ since quote monitoring is pointless. Hence the function C takes the value $C(1, 0, L)$ and we denote it $C(1)$ for simplicity.

3 Spreads and Monitoring Externalities

The results in the previous section are all conditional on a spread. In this section we determine the set of equilibrium spreads. We show that there are two important determinants of the inside spread: (1) the probability that bandits react quickly to quote updates (Φ), and (2) the number of orders submitted by a bandit (L).

3.1 The Set of Equilibrium Spreads

Consider a situation in which *all* dealers ($M_b = M \geq 2$) post the inside spread S_b^*. This is an equilibrium spread if no dealer has an incentive (i) to widen his spread or (ii) to improve upon the inside spread. The first condition requires that dealers do not expect to incur losses, that is

$$\Pi_d^*(S_b^*, M) \geq 0.$$

Let $\hat{S}(M, \Phi, L)$ be the spread such that this equation is binding (the zero expected profit spread). Using Equation (15), we get

$$\hat{S}(M, \Phi, L) = \alpha\sigma \left(\frac{Mx^s(M)C(M, \Phi, L)}{\alpha Mx^s(M)C(M, \Phi, L) + (1-\alpha)\beta\delta} \right). \tag{18}$$

In equilibrium, the inside spread must be at least $\hat{S}$ for the dealers to break even. A dealer does not improve upon the inside spread if the profit earned by posting the inside spread is at least as large as the profit he would obtain if he unilaterally undercuts. This requires

$$\Delta\Pi(S_b^*) = \Pi_d^*(S_b^*, M) - \Pi_d^*(S_b^*, 1) \geq 0. \tag{19}$$

Using Equation (15), we obtain

$$\Delta\Pi(S_b^*) = \frac{Q}{2} \left[\alpha(\sigma - S_b^*)(C(1) - x^s(M)C(M, \Phi, L)) - S_b^*((1-\alpha)\beta\delta)\frac{M-1}{M} \right]. \tag{20}$$

The dealer who undercuts gains a larger share of the order flow from liquidity traders (he executes the entire order of a liquidity trader instead of a fraction equal to $1/M$). This effect encourages undercutting and is captured by the last term inside the brackets. However, there are two counteracting effects that discourage the dealer from undercutting. First, the fraction of the dealer's depth at risk increases from $x^s(M)$ to 100%. Second, the dealer monitors more. These two effects increase the cost of market making (this is captured by $C(1) - x^s(M)C(M, \Phi, L)$). They are analyzed in detail in the following sections. For the moment, notice that if

$$C(1) - x^s(M)C(M, \Phi, L) \geq 0, \tag{21}$$

then $\Delta\Pi$ decreases with the spread. Hence the condition $\Delta\Pi \geq 0$ holds when the spread is sufficiently small. Specifically, let $\bar{S}$ be the spread such that a dealer is just indifferent between undercutting or matching the quotes ('the maximal spread'). The maximal spread solves $\Delta\Pi(\bar{S}) = 0$. Hence,

$$\bar{S}(M, \Phi, L) = \alpha\sigma \left(\frac{\Delta C}{\alpha \Delta C + (1-\alpha)\beta\delta} \right), \tag{22}$$

with

$$\Delta C(M, \Phi, L) = \frac{M\left[C(1) - x^s(M)C(M, \Phi, L)\right]}{(M-1)}. \tag{23}$$

A dealer is better off not improving upon the inside spread when $S_b^* \leq \bar{S}$. We conclude that S_b^* is an equilibrium spread if and only if it belongs to $[\hat{S}, \bar{S}]$. The next lemma provides the condition under which this interval is nonempty.

Lemma 3. *There exists an equilibrium with $M \geq 2$ dealers posting the inside spread if and only if $\hat{S}(M, \Phi, L) \leq \bar{S}(M, \Phi, L)$, that is*

$$1 - \frac{1}{M} \leq \frac{C(1) - x^s(M)C(M, \Phi, L)}{C(1)}. \tag{24}$$

The left-hand side represents the increase in the market share of a dealer who undercuts. The cost associated with undercutting is captured by the term on the right-hand side of the inequality. When Inequality (24) is strict, the maximal spread is strictly larger than the zero profit spread ($\bar{S} > \hat{S}$) and non-competitive spreads can be sustained in equilibrium. Below we study in detail the conditions under which this inequality holds.

3.2 The Effect of Monitoring Externalities

We now show how the positive externality associated with news monitoring helps dealers earn strictly positive expected profits, whereas the negative externality reduces the number of dealers who post the inside spread. In order to better convey the intuition, we assume in this section that a bandit can only submit one order of the maximum order size ($L = 1$). Analysis of the general case is deferred to Section 4.3. This means that $x^s(M) = Min\{1, L/M\} = 1/M$. In this case, Inequality (24) simplifies to

$$C(1) - C(M, \Phi, 1) \geq 0. \tag{25}$$

For $\Phi = 0$, the dealers' monitoring costs and therefore $C(M, 0, 1)$ decrease with the number of dealers posting the inside spread (see Equation (16)). We refer to this effect as the cost sharing effect. Intuitively, the number of dealers grows, each dealer can free ride on a larger number of dealers' monitoring efforts without facing an increase in the risk of being picked off. This is a result of the positive externality of monitoring which exists when Φ is sufficiently small. Accordingly, for $\Phi = 0$, Inequality (25) is always (strictly) satisfied and we obtain the following result.

Proposition 3. *In the absence of quote monitoring by the bandits ($\Phi = 0$),*

1. *All the dealers post the inside spread in equilibrium (no sidelined dealers).*
2. *There is a multiplicity of equilibrium spreads: any spread $S_b \in [\hat{S}(M, 0, 1), \bar{S}(M, 0, 1)]$ is a Nash equilibrium. For all the equilibria in which the inside spread is strictly larger than $\hat{S}(M, 0, 1)$, the dealers earn strictly positive expected profits.*[12]

The cost sharing effect deters dealers from improving upon the inside spread and explains why non-competitive spreads can be sustained. When quote monitoring is possible, $\Phi > 0$, there is a counteracting effect. Instead of simply lowering the monitoring costs, an additional dealer at the inside spread also increases the number of potential quote updates that bandits can use to learn about news. This effect (the source of the negative externality) lowers dealers' incentive to share monitoring costs. In particular the cost of market making with two dealers may then be larger than with only one dealer, despite the cost sharing effect. A second dealer at the inside enables bandits to exploit quote updates. This triggers a jump in the risk of being picked off relative to the case with one dealer and therefore matching the quotes of a single dealer can be suboptimal. This is the rationale for equilibria with only one dealer at the inside.

Lemma 4. *In the presence of quote monitoring by the bandits ($\Phi > 0$), we observe that either (a) all the dealers post the inside spread ($M_b^* = M$) or (b) only one dealer posts the inside spread ($M_b^* = 1$), in equilibrium.*

The bandits' ability to quickly exploit the information contained in quote updates (the value of Φ) determines the nature of the equilibrium as shown in the next two propositions.

Proposition 4. *There exists $\Phi^*(M, N) \in (0, 1)$ such that when $0 \leq \Phi \leq \Phi^*(M, N)$:*

1. *All the dealers post the inside spread in equilibrium.*

2. *There is a multiplicity of equilibrium spreads: any spread* $S_b \in [\hat{S}(M, \Phi, 1), \bar{S}(M, \Phi, 1)]$ *is a Nash equilibrium. Dealers earn strictly positive expected profit when* $S_b > \hat{S}(M, \Phi, 1)$.

When dealers react sufficiently quickly to quote updates ($\Phi \leq \Phi^*$), news monitoring by each dealer is a positive externality for the other dealers. Thus, the cost sharing effect dominates. For $\Phi > \Phi^*$ the increased risk of being picked off due to quote monitoring dominates the cost sharing effect. This is formally stated in the next proposition. We denote the zero profit spread when only one dealer posts the inside spread by $\hat{S}(1)$.

Proposition 5. *When* $\Phi^*(M, N) < \Phi \leq 1$, *the Nash equilibrium of the quoting stage is such that only one dealer (*$M_b^* = 1$*) posts the inside spread, which is* $S_b^* = \hat{S}(1)$. *The expected profit of the dealer posting the inside spread is zero.*

In order to prevent bandits from exploiting the quote updates, dealers undercut each other until a single dealer, who breaks even, remains at the inside spread. If another dealer were to match this inside spread, then the two dealers would incur a loss. This is due to the jump that results in the probability of being picked off. Consequently, a large relative advantage in quote monitoring for the bandits may dramatically reduce liquidity (a form of market breakdown).[13]

3.3 The Effect of Multiple Orders by Bandits

In this section we study the effect of the maximum quantity that bandits are allowed to trade, L. Recall that a dealer's participation rate in trades initiated by bandits is

$$x^s(M) = \begin{cases} 1 \quad if \quad L \geq M, \\ L/M \quad if \quad L < M. \end{cases} \tag{26}$$

Notice that $x^s(M) > \frac{1}{M}$ when $L > 1$ and $M > 1$. Now consider a dealer who undercuts his competitors and suppose $L > 1$. His participation rate in bandits' trades increases but relatively less than when $L = 1$, since $1 - x^s(M) < 1 - 1/M$. This makes undercutting more attractive when $L > 1$ since part of the cost to undercutting (greater exposure to bandits) is smaller than when

$L = 1$ whereas the benefit (greater participation rate in liquidity trades) is unchanged. This effect strengthens the dealers' incentives to price compete. Nevertheless, the cost sharing effect that we identified in the previous section still holds. It remains dominant as long as L is not too large. Proposition 4 generalize as follows.

Proposition 6. *There exists $L^*(M) > 1$ such that for $L \leq L^*(M)$ and $\Phi \leq \hat{\Phi}(M, N, L)$: (i) all the dealers post the inside spread in equilibrium and (ii) there are multiple equilibrium spreads: any spread $S_b \in [\hat{S}(M, \Phi, L), \bar{S}(M, \Phi, L)]$ is an equilibrium.*

The cut-off $\hat{\Phi}$ is the value of Φ such that Inequality (24) is binding (i.e., is such that the maximal spread and the zero profit spread are identical). For values of Φ strictly smaller than this cut-off, the zero profit spread is strictly smaller than the maximal spread and dealers earn rents in equilibria where $S_b^* > \hat{S}$. In the appendix we show that $\hat{\Phi}$ decreases with L. It is equal to Φ^* when $L = 1$ and it is equal to zero when $L = L^*(M)$. Allowing bandits to place multiple orders shrinks the set of values of Φ for which dealers can sustain non-competitive spreads. This set is always empty for $L > L^*(M)$. In this case, there is no positive value of Φ (and therefore no positive $\hat{\Phi}$) such that Inequality (24) is satisfied.

Proposition 7. *Suppose that either (a) $\Phi > \hat{\Phi}(M, N, L)$ and $L \leq L^*(M)$ or (b) $L > L^*(M)$. Then the unique equilibrium is such that only one dealer posts the inside spread which is equal to $\hat{S}(1)$ and he earns zero expected profits.*

These results generalize Proposition 5. Under the conditions of Proposition 5, the benefit to undercutting dominates the cost of undercutting and a zero expected profit equilibrium ensues. When $\Phi > \hat{\Phi}$ and $L \leq L^*(M)$, the result obtains because sharing the monitoring costs is no longer attractive, as explained in the previous section. When $L > L^*$, the result obtains because the increase in a dealer's exposure to bandits' trades is too small to deter undercutting. It is noteworthy that only one dealer posts the inside spread in equilibrium when L is large.[14] This suggests that unbridled trading by bandits can result in a decline in the total quoted depth. It also provides a justification for a limit on the number of positions that a bandit can initiate within a short interval of time. We conclude this section by considering the effect of a change in Φ on the set of equilibrium spreads.

Lemma 5. *Suppose $L \leq L^*(M)$ and $\Phi \leq \hat{\Phi}$. The zero expected profit spread, $\hat{S}$, increases with Φ whereas the maximal spread, $\bar{S}$, decreases with Φ. Furthermore $\hat{S}(M, \hat{\Phi}, L) = \bar{S}(M, \hat{\Phi}, L) = \hat{S}(1)$. In the other cases, the equilibrium spread is $\hat{S}(1)$, which does not depend on Φ.*

An increase in the bandits' relative advantage in quote monitoring increases the cost of market making (see Lemma 2). This explains why the zero expected profit spread increases with Φ. Dealers gain less in sharing monitoring costs when Φ increases and non-competitive spreads are more difficult to sustain. Thus, an increase in Φ reduces the maximal spread.

Figure 2 summarizes the results. When $\Phi \leq \hat{\Phi}$ and $L \leq L^*(M)$, there is a multiplicity of equilibrium spreads. Two equilibria are of particular interest: (1) the maximal spread equilibrium ($S_b^* = \bar{S}$), which is the preferred equilibrium for a dealer and (2) the zero expected profit equilibrium ($S_b^* = \hat{S}$), which is preferred by liquidity traders.

4 Market Design

In this section we analyze some market design issues that are motivated by some actual and proposed trading rules. We then show how theses policies affect spreads and price discovery in our model.

Nasdaq responded to the dealers' complaints about the SOES bandits by proposing to replace SOES with N*prove (1994) and NAqcess (1995). These systems were never approved by the SEC. One common feature of these trading systems is a delayed execution feature that allows dealers a short time interval during which they could decline to accept an incoming trade. This feature relaxes the firm quote requirement and consequently makes it harder for bandits to execute trades. In particular, trading strategies that rely on quote monitoring are less effective under these rules. Thus, we can interpret these proposals as a shift of the relative advantage in quote monitoring to the dealers (a lower Φ in the model).

Interestingly, one of the existing trading rules on Nasdaq also affects the dealers ability to update their quotes rapidly. Nasdaq's Autoquote Policy prohibits software that would automatically update one dealer's quotes as a function of other dealers' quotes. By forcing a dealer to update his quotes manually when he receives an alert, this policy increases the time required for him to adjust his quotes. Thus, allowing auto-quoting can also be interpreted as a shift of the relative advantage in quote monitoring favoring the dealers.

We consider a base case where Φ is positive but small enough, $\Phi \leq \hat{\Phi}$, and $1 \leq L \leq L^*$ to avoid a situation with only one dealer posting the inside spread.[15] We refer to our base case as a market design with a firm quote rule. We then compare this case with a *relaxed* quote rule where Φ is lower than in the base case, i.e., we shift the advantage in quote monitoring to the dealers. We take this lower value to be zero ($\Phi = 0$), without affecting the results.

Corollary 1. *When the equilibrium of the quoting stage is the maximal spread equilibrium (zero expected profit equilibrium), the inside spread is smaller (larger) under the firm quote rule.*

Consider Figure 2. If the dealers post the zero expected profit spread, then the equilibrium spread is clearly larger when $\Phi > 0$. This reflects the fact that the adverse selection risk is larger when bandits can use the information revealed by quote updates to pick off dealers. However if the dealers post the maximal spread, the conclusion is reversed: the equilibrium spread is smaller when $\Phi > 0$.

Corollary 2. *The monitoring level chosen by a dealer in equilibrium is always larger under the firm quote rule, both in the zero expected profit and in the maximal spread equilibria.*

A firm quote rule strengthens the dealers' incentive to be first to discover new information because it increases the likelihood that bandits (rather than dealers) benefit from quote updates. Free riding on other dealers' monitoring becomes risky.[16] One implication is that dealers' quotes will reflect new information more quickly under the firm quote rule. The speed of price discovery, however, is determined by the *aggregate* monitoring level, $\lambda_A + \gamma_A$.[17]

Corollary 3. *In the maximal spread equilibrium (zero expected profit equilibrium), the aggregate monitoring level, $\lambda_A^* + \gamma_A^*$, is larger (smaller) under the firm quote rule.*

Thus, a firm quote rule may or may not improve price discovery. On the one hand, it strengthens the dealers' incentives to monitor. On the other hand, it weakens the bandits' incentive to monitor, since they can use the free information contained in quote updates to pick off dealers. In the zero expected profit equilibrium, this effect is reinforced by the fact that the spread is larger under the firm quote rule (the bandits' monitoring effort decreases with the spread). Thus, in this case the aggregate monitoring is lower. In the maximal spread equilibrium, the spread is smaller under the firm quote rule. In this case, the increase in the dealers' aggregate monitoring level is larger than the reduction in the bandits' monitoring level, and price discovery is improved.

To sum up, our analysis provides some support for both the bandits' and the dealers' arguments. If the dealers are playing the maximal spread equilibrium, a policy that makes it easier for the bandits to pick off stale quotes may *both* improve price competition and price discovery. This vindicates the argument that a firm quote requirement provides "market discipline." On the other hand, if the dealers are posting the zero profit spread, a policy that enables bandits to pick off stale quotes would increase the spread and slow down price discovery. This finding supports the dealers' argument that the firm quote requirement, in presence of bandits, impairs market quality.

5 Testable Implications

A major question in the SOES controversy is whether or not SOES bandits cause dealers to post wider spreads. Our goal is to study this issue empirically. In this section we develop some comparative statics that we use in our empirical analysis. We first consider the impact of an increase in the number of bandits on the equilibrium spread. When there is a multiplicity of equilibria we focus on the zero expected profit and the maximal spread equilibrium.

Proposition 8. *A larger number of bandits increases the equilibrium spread, ceteris paribus.*

The intuition is as follows. In equilibrium, the probability that a bandit submits an order when two or more dealers post the inside spread is

$$\begin{aligned} \alpha\left(Prob(f \in \mathcal{N}) + \Phi Prob(f \in \mathcal{M})\right) &= \alpha\Big(\frac{\gamma_A^*}{\lambda_A^* + \gamma_A^*} + \Phi\frac{\lambda_A^*}{\lambda_A^* + \gamma_A^*}\Big) \\ &= \alpha\left(\frac{N + (\Phi - h(\Phi, L))}{N + 1 - h(\Phi, L)}\right), \end{aligned} \tag{27}$$

where the last equality follows from Proposition 2. The same expression for this probability with $\Phi = 0$ is obtained in equilibria with only one dealer. Thus, an increase in the number of bandits increase the risk of the dealers being picked off. Proposition 8 yields the following testable hypothesis.

Hypothesis 1: Stocks with a higher level of bandit activity have wider spreads, ceteris paribus.

Testing Hypothesis 1 is not straightforward because the bandit activity itself depends on the spread. We need to control for this effect. To this end, we extend the model assuming that each bandit bears a fixed entry cost, $K > 0$, that is sunk at the beginning of the trading game. This fixed cost represents, for instance, the cost of acquiring computer systems for trading. For a given spread, a bandit's expected profit (see Proposition 2) net of the fixed cost K is

$$\Pi_s^*(S_b, N) - K = \alpha Q^s(\sigma - S_b)\left[\frac{2N(N+1-h(\Phi,L)) - (N-h(\Phi,L))^2}{4N(N+1-h(\Phi,L))^2}\right] - K \tag{28}$$

The same expression obtains when a single dealer posts the inside spread, with $\Phi = 0$ in this case. Bandits take the spread as given and enter if their net expected profit is positive. Clearly, the net expected profit decreases in the number of bandits and is negative when this number is large. The number of bandits who enter, $N^*(S_b)$, is such that the net expected profit is equal to zero.[18] Note that an increase in the spread reduces N^* since a bandit's net expected profit decreases in the spread.

Proposition 9. *A larger spread leads to fewer bandits, everything else equal.*

This result gives us our second main prediction.

Hypothesis 2: Stocks with wider spreads have lower levels of bandit activity, ceteris paribus.

Hypotheses 1 and 2 underscore the interdependence between the spread and the number of bandits. Consequently, we will test Hypotheses 1 and 2 using a *simultaneous equations* framework with the spread and the level of bandit activity as endogenous variables. In order to do this we need to determine how the other model variables (σ, Q, M, δ) influence the spread and/or SOES bandit activity.

Corollary 4.

1. *For a given number of bandits, an increase in the average size of liquidity trades (δ) shrinks the spread. An increase in volatility (σ) widens the spread.*

2. *For a given spread, an increase in the minimum quoted depth (Q) or an increase in volatility (σ) triggers the entry of bandits.*

The above result for the spread is intuitive. The second part of the corollary follows since an increase in the minimum quoted depth or in the asset volatility raises bandits' expected profits, all else being equal. In our empirical analysis, we also use the number of dealers in a stock as a control variable, but we do not formulate predictions regarding the effect of the number of dealers on the spread. Actually, the model can not yield clear-cut predictions for the direction of this effect. In order to illustrate this fact, we consider a special case in the next corollary.

Corollary 5. *Suppose that $L = 1$ and $\Phi \leq \Phi^*(M, N)$. In the zero expected profit equilibrium, the spread decreases in the number of dealers posting the inside spread. In the maximal spread equilibrium, the spread can increase with the number of dealers posting the inside spread when Φ is large.*

Recall that if $\Phi \leq \Phi^*$ all the dealers post the inside spread in equilibrium and share the monitoring costs. It follows that the cost of market making and therefore the zero expected profit spread decreases with the number of dealers. At the same time, cost sharing makes undercutting less attractive when the number of dealers is large. Hence an increase in the number of dealers makes it easier to sustain non-competitive spreads. This explains why, counterintuitively, an increase in the number of dealers may result in a larger spread in the maximal spread equilibrium.

For a given spread, a decrease in the minimum quoted depth induces the entry of fewer bandits. This decline in the number of bandits reduces the risk of being picked off for the dealers and reduces the spread. Hence a change in Q *indirectly* affects the spread because it influences the number of bandits. Notice that the decrease in the spread counter-balances the initial negative impact of a reduction in Q on the number of bandits. Nevertheless, the next proposition states that despite this effect, a decrease in the minimum quoted depth reduces the number of bandits in equilibrium.

Proposition 10. *Suppose either (a) $L = 1$ and $\Phi < \Phi^*(M, 1)$ or (b) $L \geq M$. In equilibrium a reduction in the minimum quoted depth, Q, leads to (i) fewer bandits, (ii) a smaller spread, and (iii) lower level of aggregate monitoring.*

Interestingly, the minimum quoted depth has been reduced several times on Nasdaq. Nasdaq argued that this reduction would lessen SOES bandit activity and would narrow spreads. The previous proposition concurs, but it points out that a reduction in the minimum quoted depth adversely affects price discovery. Fewer bandits imply that the bandits' aggregate monitoring level decreases. Dealers also choose to monitor less, since the risk of being picked off is smaller. Eventually price discovery is impaired. The last result yields a third prediction.

Hypothesis 3: Stocks with higher minimum quoted depth have (i) larger spreads and higher levels of bandit activity.

In line with the second part of Hypothesis 3, Harris and Schultz (1997) and Barclay *et al.* (1999) find a decline in the number of trades initiated by SOES bandits after the reduction in the minimum quoted depth in 1994 and 1997, respectively.

Remark. In order to establish the last Proposition 10, we must determine how a change in Q affects (a) the spread and the number of dealers posting this spread and (b) the number of bandits, in equilibrium. This is difficult because a change in the number of bandits can trigger a shift from an equilibrium in which all the dealers post the inside spread to an equilibrium with a single dealer posting the inside spread (the cut-off $\hat{\Phi}$ depends on N). This creates discontinuities in the bandits' expected profit function when N varies. The conditions on the parameters guarantee that this technical problem does not arise. When $L = 1$ and $\Phi < \Phi^*(M, 1)$, all the dealers post the inside spread in equilibrium, independent of the number of bandits. When $L \geq M$, a single dealer posts the inside spread in equilibrium, independent of the number of bandits as well. Notice that the proposition covers all the possible situations in equilibrium.

6 Empirical Analysis

Armed with the results of the previous section, we are now able to address empirically some of the key questions in the SOES debate: Does an increase in SOES bandit activity increase the spread? Is the maximum SOES quantity an effective policy instrument for influencing SOES bandit activity?

6.1 Methodology

We need a proxy for the number of SOES bandits since we do not observe it directly. A natural measure of their activity is the unconditional probability of observing a trade initiated by a bandit. In our model, this probability is given by Equation (27) and is strictly increasing in the number of bandits. The qualitative effects of a change in the exogenous parameters on the number of bandits and this probability are identical.

But how do we identify trades initiated by bandits? Harris and Schultz (1997) show that SOES trades occurring in clusters (several maximum-size SOES trades in rapid succession) are very likely to be initiated by bandits. Accordingly, we use the probability of a SOES *cluster* as our proxy for the probability of a trade initiated by a bandit. We define a *cluster* as an uninterrupted sequence of three SOES orders of the maximum size, at the same price, within 30 seconds.[19] Our proxy is then defined as the number of SOES clusters divided by the total number of trades.

We estimate the following system of simultaneous equations for a cross-section of stocks:

$$\begin{cases} soes_i = & a_1 + a_2 spr_i + a_3 vlty_i + a_4 maxq_i + \epsilon_1 \\ spr_i = & b_1 + b_2 soes_i + b_3 vlty_i + b_4 ndlr_i + b_5 liqd_i + \epsilon_2, \end{cases} \tag{29}$$

where $i = 1, \ldots, I$ index the stocks and the variables in the equation system are: the probability of a SOES cluster (soes), the bid-ask spread (spr), the volatility of the stock returns (vlty), the maximum quantity that can be traded in SOES (maxq), the number of dealers in the stock (ndlr), and the average size of liquidity trades (liqd). We define these variables in more detail below.

The first equation determines the probability of observing a SOES *cluster* as a function of the bid-ask spread, the volatility of the asset, and the maximum SOES quantity. The second equation determines the spread as a function of the probability of a SOES *cluster*, the volatility of the asset, the number of dealers, and the average size of liquidity trades. Our two main predictions are that the effect of the spread on the bandit activity is negative, $a_2 < 0$, and that the effect of the bandit activity on the spread is positive, $b_2 > 0$. Corollary 4 provides the expected signs for the other independent variables. Recall that we can not sign the effect of the number of dealers unambiguously.

6.2 Data

Our data are provided by Nasdaq and it includes transactions and dealer quotes for December 1996. In taking the two-equation model to the data we face the following two difficulties. Previous research and anecdotal evidence suggest that bandit activity is very heavily concentrated in the large and active stocks whereas many less actively traded stocks have very little or no bandit activity.

Provided there is enough variation in the key instruments, i.e., the maximum SOES quantity and the number of dealers, we could estimate Equation (29) for a cross-section of actively traded stocks. The problem is that the rules for assigning the maximum SOES quantity imply that there is little or no variation in the maximum SOES quantity for a sample that is restricted to the very active stocks.[20] In order to address this problem we select a larger number of stocks than previous studies. The second challenge is that our dependent variable is defined as the ratio of the number of SOES cluster to the total number of trades. The normalization by the number of trades is important since our theoretical predictions concern the relative likelihood of a trade by a bandit rather than the absolute number of bandit trades. To control for this problem and to check the robustness of the results obtained for the first sample we construct a second sample. There is not much overlap between the two samples–only about 8% of the stocks in the first sample are included in the second sample.

The selection criteria used for the first sample is trading volume. Using a cut-off of four million shares for the monthly trading volume and a minimum average price of five dollars we obtain a sample of 310 stocks. Our sample includes many of the stocks that are frequently mentioned as favorites among the bandits, but also other active stocks with little or no bandit activity as measured by our proxy (see Table 3).

We construct a second sample using the following selection criteria. We rank all NASDAQ stocks with a price above three dollars by the number of trades in December 1996. We select the top one hundred stocks with a maximum SOES quantity of 500. These stocks are then matched with stocks with a maximum SOES quantity of 1000 using number of trades as the matching criteria. By selecting a fixed number of stocks with a smaller SOES size we get large variation in the maximum SOES quantity and by matching on the number of trades we ensure that the cross-sectional variation in our proxy for bandit activity is not driven by variation in the number of trades. The disadvantage of this sample and any sample of less actively traded stocks is that the overall level of bandit activity tends to be small making it harder to pinpoint the effect of changes in bandit activity.

Table 3 reports, for each of the variables we use in our analysis, the mean, median, standard deviation, minimum, and maximum. The first four rows report these statistics for the total number and frequency of SOES clusters, and the total number of SOES trades and non-SOES trades. The average number of clusters is 204 in the first sample and 16 in the second sample, suggesting that bandit activity as measured by our proxy is concentrated in the most active stocks. The median number of clusters of 61 and 2, respectively, provide evidence of a skewed distribution with a lot of bandit activity concentrated in a relatively small number of stocks. The bid-ask spread is measured as a weighted time series average of the relative inside spread. Each observation is given a weight that is proportional to the time the observed spread was in effect. The standard deviation and the range for the spread suggest that there is substantial variation in this variable both within each sample and across the two samples. On average stocks in the second sample have a bid-ask spread of 2.57% compared with an average of 1.3% for the first sample.

The volatility is measured by the standard deviation of the half-hour returns based on the mid-quotes, excluding overnight returns. The maximum SOES quantity is a discrete variable that is equal to 1000 (for 294 stocks), 500 (for 10 stocks), and 200 (for 6 stocks) in the first sample. By construction the second sample is evenly split between a SOES quantity of 1000 and 500. The number of dealers for each stock is defined as the time-series average of the number of active dealers in the stock. We compute the average trade size for all trades excluding trades that were part of a SOES cluster. Note that SOES accounts for only a small fraction of the total trading volume for most stocks. Accordingly, we find that the average trade size is larger than the maximum quantity that can be traded in SOES. The last two rows report statistics for the market capitalization and the average price. These two variables are likely to influence the bid-ask spread (see Harris (1994)), although they do not play a direct role in our model. We use them as control variables to improve the efficiency of our estimation. Overall, the companies in the second sample have a smaller market capitalization, are less actively traded by investors, and less likely to be traded by bandits.

We use transformations of some of the variables discussed above in the estimation. In the subsequent discussion our proxy for SOES bandit activity is defined as the logarithm of the odds ratio for clusters, i.e., $ln(\frac{p}{1-p})$, where p is the proportion of clusters among all trades.[21] We normalize the average trade size by the maximum SOES quantity so that the resulting variable, referred to below as the liquidity demand, corresponds to the δ in the model. Finally, we take the logarithm of the market capitalization and the average price.

Table 4 presents the correlation matrix for the variables that we use in the estimation. Notice that the correlation between the average bid-ask spread (spr) and the proxy for SOES bandit activity (soes) is -0.684 and -0.507, respectively, in the two samples. This negative correlation is consistent with the observation that more bandits are active in stocks with smaller spreads (Proposition 9). This does not rule out that an increase in bandit activity, *holding everything else equal*, leads to wider spreads as predicted by Proposition 8.

6.3 Empirical Results

Table 5 reports the parameter estimates and corresponding p-values for our two-equation model (Equation (29)).[22] The estimates for the endogenous variables provide mixed support for the predictions of the model. The parameter estimate for the bid-ask spread in the SOES Equation is negative, with a p-value less than 0.001, for both sample. This means that an increase in the spread is an effective defense against trading by bandits. On the other hand, we find only limited support for the dealers' claim that trading by the bandits forced them to widen their spreads: in the Spread Equation, the coefficient on bandit activity is positive in both samples. The effect of bandit on the spread is statistically weak, however, with the coefficient significant only at the 10% level (p-value of 0.079) in the first sample. In the second sample the coefficient is not significantly different from zero. Possible explanations for this finding are discussed in the next section.

The following numerical example illustrates the economic significance of these parameter estimates for an average stock in the first sample. Consider a stock with an average *probability* of a SOES cluster, which corresponds to 1.297%. A one standard deviation increase in this probability (roughly 128 basis points) leads to an increase in the bid-ask spread of 30 basis points (which corresponds to a 0.44 standard deviation increase).[23] On the other hand, a one standard deviation increase in the spread, which roughly corresponds to 68 basis points, leads to a 81 basis points drop in the probability of a SOES cluster for an average stock (this corresponds to a 0.63 standard deviation decrease).

The estimated coefficients for the maximum SOES quantity in the SOES Equation are positive and highly significant for both samples. The coefficients on volatility are both positive, but only the coefficient in the first sample is estimated precisely. In the Spread Equation the coefficient on volatility is positive (p-values of 0.058 and 0.115, respectively). All the estimates above have the predicted signs. In line with intuition, the coefficient on the number of dealers is negative in both samples with p-values less than 0.001. The trade size does not appear to play an important role in determining the spread in the first sample; the coefficient has a p-value of 0.453 whereas the coefficient on trade size is negative and significant, as predicted by the model, in the second sample.

Each estimated parameter in Table 5 measures the impact on the spread (or bandit activity) of one exogenous variable, holding all other variables constant. In order to study how a change in the maximum SOES quantity Q would indirectly affect the spread, we estimate two "reduced-form" regressions. Table 6 report the results for these regressions of the endogenous variables on all the exogenous variables.

In Table 6, the coefficient on the maximum SOES quantity, in the Spread Equation, is positive (with a p-value of 0.058 and 0.213, respectively). This implies, other things equal, that stocks with a lower minimum quoted quantity have tighter spreads, as predicted by Proposition 10. According to our model, the effect of the maximum SOES quantity on the spread is indirect: An increase in this variable attracts bandit activity, which in turn tends to increase the spread. Hence, the low statistical significance is consistent with our previous finding that SOES bandit activity has only a moderate impact or, as is the case for the second sample, no impact on the spread. A back of the envelope calculation shows that a change in the maximum SOES quantity from 500 to 1000 shares would increase the spread by 19 basis points (or a 0.28 standard deviation increase) for stocks in the first sample.

In the SOES Equation we also find a positive and significant (p-value < 0.001) coefficient on the maximum SOES quantity in both samples. The coefficient estimate of 0.0011604 for the first sample implies that increasing the maximum SOES quantity from 500 to 1000 shares leads to an increase in the probability of a SOES cluster of roughly 100 basis points which corresponds to a 0.78 standard deviation increase.

Notice that the coefficient on the number of dealers is positive in the SOES Equation. Stocks with a higher number of dealers have lower spreads, which would tend to attract more bandits. Bandits may also focus on stocks with a large number of dealers because stale quotes occur more frequently in such stocks.

In the reduced form regressions (Table 6), volatility has a positive impact on bandit activity. The effect is not statistically different from zero, however (p-value of 0.259 and 0.376, respectively). Recall that in the reduced-form regressions, we do not control for the effect of the spread on the number of bandits. It turns out that volatility has a positive impact on the spread. Hence the coefficient on volatility reflects a direct positive effect of volatility on bandit activity (confirmed in Table 5) and an indirect negative effect via the spread. Our empirical results suggest that the two effects essentially cancel so that volatility does not significantly affect the bandit activity.

6.4 Summary and Discussion

Overall, our results are consistent with a market where the extent of trading by the bandits is strongly influenced by variables that predict profitability: (i) the bid-ask spread, (ii) the maximum SOES quantity. However, our empirical results provide very weak evidence in support of the hypothesis that increased bandit activity leads to wider spreads. This suggests that the dealers' trading costs or at least the bid-ask spreads are not very sensitive to losses due to bandit trading. The findings of Harris and Schultz (1997) also support this conclusion.[24]

At first glance our result may seem to be at odds with evidence of positive bandit profits as reported in Harris and Schultz (1998). The result is also puzzling given the time and resources that dealers have spent lobbying against the bandits. It is, of course, important to realize that the documented bandit profits concern a relatively small number of very active stocks whereas our results for the bid-ask spread are obtained for broader cross-sections of stocks. It is possible that on average the effect of the bandits on the spread is too small to detect even if there was a stronger effect in a smaller subset of stocks. The marginally significant effect found for the first sample and the insignificant effect found for the second sample are consistent with this argument. Below we will discuss some alternative explanations for our findings.

Several institutional rules may make it difficult to measure the impact of bandit activity on spreads. First, in our sample period, the minimum price increment was \$1/8 for most stocks. For some stocks, this may be larger than the compensation required by dealers for the risk of being picked off by bandits. In this case, an increase in bandit activity will have no discernible impact on observed spreads even if it increases the cost of market making. Second, many larger trades receive price improvements. In our model, dealers compensate the losses inflicted by bandits by quoting larger spreads. In reality, they may decide to leave their quoted spread unchanged but to offer price improvements less frequently. In order to examine this explanation our model would need to be extended to allow the dealers a richer set of choices. This analysis is beyond the scope of this paper.[25]

To sum up, given the above difficulties one should not conclude that our results suggest that a very high levels of SOES bandit activity does not affect the trading costs of a stock. What our results suggest is that for a typical stock in this market, or at least in our relatively large cross-section, SOES bandit trading level is not an important determinant of the trading costs.

7 Conclusion

We develop a model of information monitoring and market making in a dealership market. Our analysis is motivated by the controversy concerning SOES bandits on Nasdaq, but can be viewed more broadly as well. Dealers choose to invest in costly information monitoring in order to reduce the risk of being picked off. By matching the quotes of other dealers rather than undercutting, dealers can share the monitoring costs. When active traders such as the SOES bandits can use the information revealed by quote updates to pick off dealers they add competitive pressure and force dealers to quote narrower spreads and quickly update their quotes. On the other hand, when this picking off risk becomes to large dealers may refuse to post quotes and we observe a dramatic decrease in liquidity. Thus, unbridled trading by SOES bandits or other active traders may harm market liquidity, as the opponents of the SOES bandits have argued.

Important changes in Nasdaq trading rules have been implemented following the period we study. Trading in SOES has decreased following the introduction of the order handling and actual size rules in 1997 according to Barclay et.al. (1999). Based on our model we would expect a decrease in activity because a smaller minimum quoted depth, a consequence of the actual size rule, tend to decrease bandit profits. More recently, Nasdaq's new SuperMontage system includes an updated version of the SOES system called SuperSoes. SuperSoes retains the key feature of the old system namely automatic execution. One important difference is that dealers can use the new system for both agency and proprietary orders. It therefore creates a level playing field, something that dealers have called for. The equal access feature in SuperSoes makes it similar to other automatic trading systems such as electronic limit order markets. In an electronic limit order market any market participant can make a market by placing a limit order or trade against limit orders placed by other traders. Monitoring of public information is useful for two reasons. First, a trader can reduce the risk of her order being picked off. Second, she can increase the chance of observing and picking off other traders' stale orders. A natural question for future research is to sort out how price discovery and liquidity provision is affected when all traders can play the roles of the dealers and the bandits.

Proofs

Proof of Proposition 1. Using Equation (7)

$$\frac{\partial \Pi_d(\lambda_i, \lambda_{-i}, \gamma)}{\partial \lambda_m} = -\alpha \left[x^s(M_b) \frac{\partial Prob(f \in \mathcal{N})}{\partial \lambda_m} + x^s(M_b - 1)\Phi \frac{\partial Prob(f \in \mathcal{M}_b \backslash i)}{\partial \lambda_m} \right] \frac{(\sigma - S_b)Q}{2}$$

$$= \frac{\alpha}{(\lambda_A + \gamma_A)^2} \left[x^s(M_b)\gamma_A - \Phi x^s(M_b - 1)(\gamma_A + \lambda_i) \right] \frac{(\sigma - S_b)Q}{2} \quad \forall m \neq i. \tag{A1}$$

Equation (A1) is positive if and only if $x^s(M_b)\gamma_A - \Phi x^s(M_b - 1)(\gamma_A + \lambda_i) \geq 0$ and $\bar{\Phi}$ follows directly. Since $x^s(M_b)$ (weakly) decreases with M_b, $\bar{\Phi} < 1$.

Proof of Lemma 1. Suppose (to be contradicted) that there exists a Nash equilibrium in which dealer i chooses λ_i and dealer i' chooses λ_i^* and $\lambda_i^* > \lambda_{i'}^*$. The difference of the first order conditions (Equation (8)) for i and i' yield

$$\frac{\alpha(\sigma - S_b)Q}{2(\lambda_A + \gamma_A)^2} \left[(\Phi x^s(M_b - 1))(\lambda_{i'}^* - \lambda_i^*) \right] = \Psi'(\lambda_i^*) - \Psi'(\lambda_{i'}^*). \tag{A2}$$

Since $\lambda_i^* > \lambda_{i'}^*$, the L.H.S of this equality is strictly negative. But since $\Psi'(.)$ is increasing, the right hand side is strictly positive. A contradiction. A similar argument applies to the bandits.

Proof of Proposition 2. First we note that Equation (11) can be rewritten as

$$\frac{\alpha(\sigma - S_b)Q^s(M_b)}{M_b(M_b\lambda^* + N\gamma^*)^2} \left[N\gamma^* + h(\Phi, L)M_b\lambda^* \right] = c\lambda^*. \tag{A3}$$

Thus, dividing Equation (A3) by Equation (12), we find that λ^* and γ^* must satisfy

$$\frac{N\gamma^* + hM_b\lambda^*}{(\frac{N-h}{N})M_b\lambda^* + (N-1)\gamma^*} = (\frac{M_b\lambda^*}{\gamma^*}).$$

Writing this equation in term of one unknown variable, $\Upsilon \equiv \frac{M_b\lambda^*}{\gamma^*}$, and noting that the monitoring levels must be positive ($\Upsilon \geq 0$) we find a single solution: $\Upsilon = \frac{N}{(N-h)}$. Substituting $(\Upsilon\gamma^*)$ for $(M_b\lambda^*)$ in Equation (12), we find that γ^* solves

$$\frac{\alpha Q^s(\sigma - S_b)(N-h)^2}{N(N+1-h)^2} = c(\gamma^*)^2. \tag{A4}$$

There is a unique positive solution to this equation, which yields γ^*. We then obtain λ^* using the fact that $\lambda^* = \frac{\Upsilon\gamma^*}{M_b}$. As λ^* and γ^* are uniquely defined, there is a unique Nash equilibrium in the monitoring stage. Substituting the expressions for λ^* and γ^* in Equations (7) and (9) yield Π_d^* and Π_s^*.

Proof of Lemma 2. By definition $h(\Phi, L) = \Phi\frac{Q^s(M-1)}{Q^s(M)}$. Hence $h(\Phi, L)$ increases with Φ. Using this fact and Equation (16), we get $\frac{\partial C(M,\Phi,L)}{\partial \Phi} > 0$.

Proof of Lemma 3. An equilibrium with M dealers posting the inside spread exists if and only if $\hat{S}(M, \Phi, L) \leq \bar{S}(M, \Phi, L)$. Using Equations (18) and (22), we obtain that this inequality is satisfied if and only if $Mx^s(M)C(M, \Phi, L) \leq \Delta C(M, \Phi, L)$, that is (using Equation (23))

$$\frac{x^s(M)C(M, \Phi, L)}{C(1)} \leq \frac{1}{M}. \tag{A5}$$

This yields Inequality (24) after a straightforward manipulation.

Proof of Proposition 3. Part 1. Notice that $C(M_b, 0, 1)$ decreases with M_b. Therefore, using Equation (15) with $x^s(M_b) = 1/M_b$, if

$$\Pi_d^*(S_b, M_b) = \frac{Q}{2M_b}[-\alpha(\sigma - S_b)C(M_b, 0, 1) + (1 - \alpha)\beta\delta S_b] \geq 0,$$

then

$$\Pi_d^*(S_b, M_b + 1) = \frac{Q}{2(M_b + 1)}[-\alpha(\sigma - S_b)C(M_b + 1, 0, 1) + (1 - \alpha)\beta\delta S_b] > 0.$$

Thus, a sidelined dealer is always better off matching the inside spread; an equilibrium in which a subset of dealers are sidelined when $\Phi = 0$ and $L = 1$ does not exist.

Part 2. Since $C(M, 0, 1)$ decreases with M and $C(1) = C(1, 0, L)$ (by definition), Inequality (25) is satisfied. The second part of the proposition follows.

Proof of Lemma 4. Because $L = 1$, $Q^s(M - 1) = Q^s(M) = Q$. It is then immediate that $h(\Phi, L) = \Phi$. Using Equation (16), we deduce that $C(M, \Phi, 1)$ decreases with M for $M \geq 2$. Therefore we can proceed as in the proof of Proposition 3 (1st part) to show that there is no equilibrium in which a subset of two or more dealers post the spread and some dealers are sidelined. We cannot, however, rule out the possibility that $C(1) < C(2, \Phi, 1)$ since C increases with Φ (recall that by definition $C(1) = C(1, 0, L)$).

Proof of Proposition 4. Let $\Phi^*(M, N)$ be the value of Φ such that Inequality (25) is binding, i.e. such that $C(M, \Phi, 1) - C(1) = 0$. Since $C(M, \Phi, 1)$ increases with Φ, $C(M, \Phi, 1) \leq C(1)$ if $\Phi \leq \Phi^*$ and $C(M, \Phi, 1) > C(1)$ otherwise. Thus Inequality (25) is satisfied if and only if $\Phi \leq \Phi^*(M, N)$. Using equation (16), we find that

$$C(M, \Phi, 1) - C(1) = 0 \Leftrightarrow \frac{(N+1)^2}{M} - (1 + N - \Phi)\left[(1+N)(1-2\Phi) - \Phi\right] = 0. \tag{A6}$$

This equation has only one solution in $[0, 1]$ which is

$$\Phi^*(M, N) = \frac{(1+N)(2+N)}{3+2N}[1 - \sqrt{1 - \frac{(M-1)(3+2N)}{M(2+N)^2}}] > 0, \quad for \quad M \geq 2.$$

Note that $\Phi^*(.,.)$ increases with M and N and is always less than $1/2$.

Proof of Proposition 5. When $\Phi > \Phi^*(M, N)$, Inequality (25) does not hold, and there is no equilibrium in which all the dealers pool on the inside spread. If an equilibrium exists, it must therefore feature a single dealer (an implication of Lemma 4). Denote the zero expected profit spread with a single dealer at the inside by $\hat{S}(1)$, $\hat{S}(1)$ is equal to $\hat{S}(1, 0, 1)$ in Equation (18). Zero expected profits imply

$$\Pi_d^*(\hat{S}(1), 1) = 0.$$

There are three necessary conditions. First, dealer m must make positive expected profits:

$$\Pi_d^*(S_b^*, 1) \geq 0,$$

which implies $S_b^* \geq \hat{S}(1)$. Moreover, the spread posted by the sidelined dealers must be just slightly greater than the inside spread, otherwise dealer m would widen his spread. Second, a sidelined dealer should not be better off undercutting the inside spread. This condition requires that dealer m obtains *zero expected profit*, i.e., $S_b^* = \hat{S}(1)$. Third, no sidelined dealer should be better off pooling on the inside spread with dealer m: $\hat{S}(2, \Phi, 1) > \hat{S}(1)$, that is $C(2, \Phi, 1) > C(1)$.

We show that this is the case. Recall that $\Phi^*(M, N)$ increases with M. Hence if $\Phi > \Phi^*(M, N)$ then $\Phi > \Phi^*(2, N)$. Now recall that, by definition, $\Phi^*(2, N)$ is such that $C(2, \Phi^*(2, N), 1) = C(1)$. As $C(2, \Phi, 1)$ increases with Φ, it follows that $C(2, \Phi, 1) > C(1)$ since $\Phi > \Phi^*(2, N)$.

Proof of Proposition 6. We define $F(M,\Phi,L) = Mx^s(M)C(M,\Phi,L)$. Consider a situation in which all the dealers post a spread $S_b \in [\hat{S}(M,\Phi,L), \bar{S}(M,\Phi,L)]$. Recall that this situation is an equilibrium if and only if Inequality (24) holds. This inequality can be written

$$F(L,M,\Phi) \leq C(1) \quad for \quad M \geq 2. \tag{A7}$$

Note that $C(1)$ does not depend on L, Φ and M. The function $C(M,\Phi,L)$ depends on L through $h(\Phi,L)$ (see Equation (16)). For $M \geq 2$,

$$h(\Phi,L) = \begin{cases} \Phi \quad if \quad L \leq M-1, \\ \Phi\left(\frac{M-1}{L}\right) \quad if \quad (M-1) < L < M, \\ \Phi\left(\frac{M-1}{M}\right) \quad if \quad L \geq M. \end{cases}$$

Moreover we observe that (1) $F(M,\Phi,L) = LC(M,\Phi,L)$ for $L \leq M$ and that (2) $F(M,\Phi,L) = MC(M,\Phi,L)$ is independent of L for $L \geq M$. Differentiating and using the above observations give

$$\begin{aligned} &\tfrac{\partial F}{\partial L} = C(M,\Phi,L) > 0 \quad for \quad L \leq (M-1) \\ &\tfrac{\partial F}{\partial L} = C(M,\Phi,L)\left[1 - \tfrac{(M-1)}{L(N+1-h(\Phi))}\right] > 0 \quad for \quad (M-1) < L \leq M. \end{aligned}$$

Hence $F(M,\Phi,L)$ increases with L. It is immediate that F increases with Φ since C increases with this parameter.

For each L, we define $\hat{\Phi}(M,N,L)$ as the value of Φ such that Inequality (A7) is binding (if such a value exists). Thus, for $L = 1$, $\hat{\Phi} = \Phi^*$ (see the proof of Proposition 4). Furthermore since F increases with L and Φ, it is immediate that $\hat{\Phi}$ decreases with L. We define $L^*(M)$ as the value of L such that $\hat{\Phi}(M,N,L^*) = 0$. Observe that $L^*(M) > 1$ since $\hat{\Phi}(M,N,1) = \Phi^* > 0$. For values of L larger than $L^*(M)$, there is no positive value of Φ such that Inequality (A7) can be satisfied.

Proof of Proposition 7. It is immediate that $F(M,\Phi,M) > C(1), \quad \forall\Phi$. Thus, $L^*(M) < M$. It follows that for $L \leq L^*(M)$, $F(M,\Phi,L) = LC(M,\Phi,L)$ so that F decreases with M. Hence $\hat{\Phi}$ increases with M. In turn this implies that L^* increases with M since $\hat{\Phi}$ decreases with L.

From Proposition 6, we know that an equilibrium with $M_b \geq 2$ dealers posting the inside spread exist if and only if $\Phi \leq \hat{\Phi}(M_b,N,L)$ and $L \leq L^*(M_b)$. If these conditions are not satisfied for

$M_b = M$, they can not hold for $2 \leq M_b < M$. Actually, in the proof of Proposition 6, we have shown that $\hat{\Phi}$ and L^* increases with the number of dealers. It follows that

$$\Phi > \hat{\Phi}(M,N,L) \Rightarrow \Phi > \hat{\Phi}(M_b,N,L) \quad for \quad 2 \leq M_b \leq M,$$

and that

$$L > L^*(M) \Rightarrow L > L^*(M_b) \quad for \quad 2 \leq M_b \leq M,$$

Consequently, if $\Phi > \hat{\Phi}(M,N,L)$ or $L > L^*(M)$, there is no equilibrium with $M_b \geq 2$ dealers posting the inside spread. The proof of the existence of an equilibrium with a single dealer posting the inside spread follows the steps of the proof of Proposition 5.

Proof of Lemma 5. Observe that $\hat{S}$ increases with C. Since C increases with Φ it follows that $\hat{S}$ increases with Φ. Using Equation (23), we get

$$\frac{\partial \Delta C(M,\Phi,L)}{\partial \Phi} = -\frac{Mx^s(M)}{M-1}\frac{\partial C(M,\Phi,L)}{\partial \Phi} < 0.$$

This means that ΔC decreases with Φ. As $\bar{S}$ increases with ΔC, we obtain that $\bar{S}$ decreases with Φ. By definition, $\hat{\Phi}$ is such that

$$Mx^s(M)C(M,\hat{\Phi},L) = C(1) \equiv C(1,0,L).$$

Using this remark, Equations (18) and (22) and the fact that $\hat{S}(1) \equiv \hat{S}(1,0,1)$, we deduce that

$$\hat{S}(M,\hat{\Phi},L) = \bar{S}(M,\hat{\Phi},L) = \hat{S}(1).$$

Proof of Corollary 1. Immediate using Lemma 5.

Proof of Corollary 2. Since $\Phi \leq \hat{\Phi}(M,N,L)$, all the dealers post the inside spread in equilibrium and the set of equilibrium spreads is $[\hat{S}(M,\Phi,L), \bar{S}(M,\Phi,L)]$. Since $M_b = M$, Proposition 2 yields

$$\lambda^*(\Phi) = \sqrt{\frac{N\alpha Q^s(\sigma - S_b^*)}{cM^2(1+N-h(\Phi,L))^2}}.$$

In the zero expected profit equilibrium, $S_b^* = \hat{S}(M,\Phi,L)$. Substituting $\hat{S}(M,\Phi,L)$ in the previous equation and using the expression for $\hat{S}$ (given by Equation (18)), we obtain

$$\lambda^*(\Phi) = \sqrt{\frac{N\alpha Q^s((1-\alpha)\sigma\beta\delta)}{cM^2[\alpha M x^s(M)C(M,\Phi,L) + (1-\alpha)\beta\delta](1+N-h(\Phi,L))^2}}. \tag{A8}$$

Substituting $C(M,\Phi,L)$ by its expression (given by Equation (16)), λ^* can be written as

$$\lambda^*(\Phi) = \sqrt{\frac{N\alpha Q^s((1-\alpha)\sigma\beta\delta)}{cM^2\left[\alpha N\left(Mx^s(M)(1+N-h(\Phi,L)) + \frac{x^s(M)}{2}\right) + ((1-\alpha)\beta\delta)(1+N-h(\Phi,L))^2\right]}}.$$

Recall that $h(\Phi,L)$ increases with Φ. It follows that $\frac{\partial\lambda^*}{\partial\Phi} > 0$. In the maximal spread equilibrium, λ^* is given by Equation (A8) but $C(M,\Phi,L)$ is replaced by $\Delta C(M,\Phi,L)$. As ΔC decreases with Φ, it is direct that dealers' monitoring level increases with Φ. Thus, independently of the equilibrium we consider in the quoting stage, we obtain

$$\lambda^*(0) < \lambda^*(\Phi) \quad \forall\Phi \leq \hat{\Phi}(M,N,L). \tag{A9}$$

Proof of Corollary 3. Using Proposition 2, we obtain that the aggregate monitoring level is

$$\lambda_A^*(\Phi) + \gamma_A^*(\Phi) = M\lambda^* + N\gamma^* = \sqrt{\frac{N\alpha Q^s(M)(\sigma - S_b^*)}{c}}. \tag{A10}$$

In the zero expected profit equilibrium, $S_b^* = \hat{S}(M,\Phi,L)$. Since $\hat{S}(M,0,L) < \hat{S}(M,\Phi,L)$, we obtain (using Equation (A10))

$$\lambda_A^*(\Phi) + \gamma_A^*(\Phi) < \lambda_A^*(0) + \gamma_A^*(0) \quad \forall\Phi \leq \hat{\Phi}(M,N,L).$$

Now consider the maximal spread equilibrium. In this case, $S_b^* = \bar{S}(M,\Phi,L)$. Since $\bar{S}(M,0,L) > \bar{S}(M,\Phi,L)$, we obtain

$$\lambda_A^*(\Phi) + \gamma_A^*(\Phi) > \lambda_A^*(0) + \gamma_A^*(0) \quad \forall\Phi \leq \hat{\Phi}(M,N,L).$$

Proof of Proposition 8. There are three different cases in equilibrium: (1) all the dealers post the zero expected profit spread $\hat{S}(M,\Phi,L)$; (2) a single dealer posts the zero expected profit spread

$\hat{S}(1)$; (3) all the dealers post the maximal spread $\bar{S}(M, \Phi, L)$.

Case 1. This case requires $\Phi \leq \hat{\Phi}$. Observe that $\hat{S}$ increases with C. Using Equation (16), we obtain

$$\frac{\partial C}{\partial N} = \frac{(1 - h(\Phi, L))}{(N + 1 - h(\Phi, L))^2} + \frac{(1 - h - N)}{2M(N + 1 - h)^3} \quad \forall \Phi, \forall L. \tag{A11}$$

Note that $h(\Phi, L) \leq \Phi \leq \hat{\Phi}$ and $\hat{\Phi}(M, N, L) \leq \Phi^* < \frac{1}{2}$. Hence $h(\Phi, L) < 1/2$. Using this remark, we obtain $\frac{\partial C}{\partial N} > 0$, which implies that $\frac{\partial \hat{S}}{\partial N} > 0$.

Case 2. We have

$$\hat{S}(1) = \alpha\sigma(\frac{C(1)}{\alpha C(1) + (1 - \alpha)\beta\delta}). \tag{A12}$$

By definition $C(1) = C(1, 0, L)$. Using Equation (16), we deduce that $C(1)$ increases with N. Consequently $\hat{S}(1)$ increases with N.

Case 3. Observe that $\bar{S}$ increases with ΔC. Using Equation (23), we obtain

$$\frac{\partial \Delta C}{\partial N} = \left(\frac{M}{M - 1}\right)\left(\frac{\partial C(1)}{\partial N} - x^s(M)\frac{\partial C(M, \Phi, L)}{\partial N}\right).$$

Using Equation (A11), we obtain that $\frac{\partial^2 C(M,\Phi,L)}{\partial\Phi\partial N} < 0$ for all values of Φ and L. This means that $\frac{\partial C}{\partial N}$ decreases with Φ. Since $C(1) = C(1, 0, L)$, we deduce that

$$\frac{\partial C(1)}{\partial N} > \frac{\partial C(M, \Phi, L)}{\partial N}.$$

Since $x^s(M) \leq 1$, we conclude that $\frac{\partial \Delta C}{\partial N} > 0$. Consequently $\bar{S}(M, \Phi, L)$ increases with N.

Proof of Proposition 9. Immediate using Equation (28).

Proof of Corollary 4.

Part 1. Consider the case in which the dealers post the zero expected profit spread, $\hat{S}(M, \Phi, L)$. It is immediate from Equation (18) that $\hat{S}$ increases with σ and that $\hat{S}$ decreases with δ. The argument is identical when the dealers post the maximal spread (using Equation (22)). The last possibility is that a single dealer posts a spread equal to $\hat{S}(1)$. By definition $\hat{S}(1) = \hat{S}(1, 0, L)$ which increases with σ and decreases with δ.

Part 2. Consider an increase in Q. It shifts bandits' net expected profit upward for a given value of N (see Equation (28)). This induces entry of more bandits. The effect of σ is identical.

Proof of Corollary 5. Under the assumptions on the parameters, all the dealers post the spread in equilibrium. Suppose first that they post the zero expected profit spread

$$\hat{S}(M, \Phi, 1) = \alpha\sigma\left(\frac{C(M, \Phi, 1)}{C(M, \Phi, 1) + (1-\alpha)\beta\delta}\right).$$

Using Equation (16), we obtain that $C(M, \Phi, 1)$ decreases with M. It follows that $\hat{S}(M, \Phi, 1)$ decreases with M. Now suppose that the dealers post the maximal spread

$$\bar{S}(M, \Phi, 1) = \alpha\sigma\left(\frac{\Delta C(M, \Phi, 1)}{\Delta C(M, \Phi, 1) + (1-\alpha)\beta\delta}\right).$$

Observe that it increases with ${}^{\Delta}C$. Computations yield

$$\frac{\partial \Delta C(M, \Phi, 1)}{\partial M} = \frac{1}{(M-1)^2}\left[C(M, \Phi, 1) - C(1) + \frac{N(M-1)}{2M^2(1+N-\Phi)^2}\right].$$

The term in brackets increases with Φ. It is strictly negative for $\Phi = 0$ and strictly positive for $\Phi = \Phi^*(M, N)$ (because by the definition of Φ^*, we have $C(1) = C(M, \Phi^*, 1)$). Thus there exists $\Phi' \in (0, \Phi^*)$ such that $\frac{\partial \Delta C(M, \Phi', 1)}{\partial M} = 0$. For $\Phi < \Phi'$, $\frac{\partial \Delta C(M, \Phi, 1)}{\partial M} < 0$ and for $\Phi > \Phi'$, $\frac{\partial \Delta C(M, \Phi, 1)}{\partial M} > 0$.

Proof of Proposition 10.

The Spread and the number of bandits

Case 1. $L = 1$ and $\Phi \leq \Phi^*(M, 1)$. Under these conditions, all the dealers post the inside spread in equilibrium, for all values of N (because Φ^* increases with N). An equilibrium is a pair $\{S_b^*, N^*\}$ such that (i) $\Pi_s^*(S_b^*, N^*) = K$ and (ii) $S_b^* \in [\hat{S}(M, \Phi, 1), \bar{S}(M, \Phi, 1)]$. Suppose first that the dealers post the zero expected profit spread, and substitute S_b^* by $\hat{S}(M, \Phi, 1)$ in Π_s^*. Using Equation (28), we obtain that N^* must satisfy

$$\left(\frac{\alpha(1-\alpha)\beta\delta Q\sigma}{\alpha C(M, \Phi, 1) + (1-\alpha)\beta\delta}\right)\left[\frac{2N^*(N^*+1-\Phi) - (N^*-\Phi)^2)}{4N^*(N^*+1-\Phi)^2}\right] = K. \tag{A13}$$

Other things equal, the left hand side of this equation increases with Q and decreases with N^* (because the term in bracket decreases with N^* and $C(M, \Phi, 1)$ increases with N). We deduce that when Q increases, N^* increases as well. Since $\hat{S}(M, \Phi, 1)$ increases with the number of bandits (Proposition 8), we conclude that the spread increases with Q.

Suppose now that the dealers post the maximal spread equilibrium, $S_b^* = \bar{S}(M, \Phi, 1)$. We can follows exactly the same steps as for $S_b^* = \hat{S}$. The only difference is that ΔC replaces C in the denominator of Equation (A13). But, since ΔC increases with N, the same argument applies.

Case 2. $\underline{L > M}$. Under this condition, $L > L^*(M)$ for all values of N since $L^*(M)$ is always smaller than M (see the proof of Proposition 6). In this case, a single dealer posts the inside spread in equilibrium. This spread is $\hat{S}(1)$. Then the argument is identical to *Case 1* with $S_b^* = \hat{S}(1)$.

Aggregate Monitoring

Using Proposition 2, we obtain that

$$\lambda_A^* + \gamma_A^* = \sqrt{\frac{\alpha N^* Q^s(M_b)(\sigma - S_b^*)}{c}}. \tag{A14}$$

The number of bandits in equilibrium is such that each bandit's expected profit is zero in equilibrium. Hence, using Equation (28), we obtain:

$$\alpha Q^s(M_b)(\sigma - S_b^*) = K\left[\frac{4N^*(N^* + 1 - h(\Phi, L))^2}{2N^*(N^* + 1 - h(\Phi, L)) - (N^* - h(\Phi, L))^2}\right]. \tag{A15}$$

Substituting this expression in Equation (A14) yields

$$\lambda_A^* + \gamma_A^* = \sqrt{\frac{K}{c}}\sqrt{\left[\frac{4(N^*)^2(N^* + 1 - h(\Phi, L))^2}{2N^*(N^* + 1 - h(\Phi, L)) - (N^* - h(\Phi, L))^2}\right]}.$$

The term in brackets increases in N^*. Thus, $\lambda_A^* + \gamma_A^*$ increases with N^*. Since N^* increases in Q, $\lambda_A^* + \gamma_A^*$ increases with Q as well.

References

Barclay, M., W. Christie, J. Harris, E. Kandel, and P. Schultz, 1999, "The Effect of Market Reform on the Trading Costs and Depths of Nasdaq Stocks," *Journal of Finance*, 54, 1-34.

Battalio, R., B. Hatch, and R. Jennings, 1997, "SOES Trading and Market Volatility," *Journal of Financial and Quantitative Analysis,* 32, 225-238.

Bernhardt, D., V. Dvoracek, E. Hughson, and I. Werner, "Why Do Large Orders Receive Discounts on the London Stock Exchange?," working paper, University of Colorado.

Copeland, T., and D. Galai, 1983, "Information Effects on the Bid-Ask Spread," *Journal of Finance*, 38, 1457-1469.

Dennert, J., 1993, "Price Competition between Market Makers," *Review of Economic Studies*, 60, 735-751.

General Accounting Office, "The Effects of SOES on the Nasdaq Market," United States General Accounting Office Report 98-194.

Harris, J., and P. Schultz, 1997, "The Importance of Firm Quotes and Rapid Executions: Evidence from the January 1994 SOES Rules Change," *Journal of Financial Economics*, 45, 135-166.

Harris, J., and P. Schultz, 1998, "The Trading Profits of SOES Bandits," *Journal of Financial Economics*, 50, 39-62.

Harris, L., 1994, "Minimum Price Variations, Discrete Bid-Ask Spreads, and Quotation Sizes," *Review of Financial Studies*, 7, 149-178.

Hinden, S., 1994, " Nasdaq's Big Guns Send Trading Bandits Packing," Washington Post, February

7, p. F33.

Houtkin, H., 1998, Secrets of the SOES Bandit, McGraw-Hill, Hightstown, N.J.

Kandel, E., and L. Marx, 1997, "Nasdaq Market Structure and Spread Patterns," *Journal of Financial Economics*, 35, 61-90.

Kandel, E., and L. Marx, 1999, "Odd-eight Avoidance as a Defense Against SOES Bandits," *Journal of Financial Economics*, 51, 85-102.

Kumar, P. and D. Seppi, 1994, "Information and Index Arbitrage," *Journal of Business*, 67, 481-509.

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

Table 1
Dealer i's Payoffs

Innovation in asset value	Outcome of the trading game: First trader to observe news	Second trader(s) to observe news	Probability of outcome	Dealer i's trading payoff
Good news	A bandit	All dealers	$\frac{\alpha}{2}Prob(f \in \mathcal{N})$	$-x^s(M_b)Q\frac{(\sigma-S_b)}{2}$
$v_1 = v_0 + \frac{\sigma}{2}$	Dealer i	All other dealers	$\frac{\alpha}{2}Prob(f = i)$	0
	Any dealer k, $k \neq i$	A bandit	$\frac{\alpha}{2}\Phi Prob(f \in \mathcal{M}_b \backslash i)$	$-x^s(M_b - 1)Q\frac{(\sigma-S_b)}{2}$
	Any dealer k, $k \neq i$	All other dealers	$(1-\Phi)\frac{\alpha}{2}Prob(f \in \mathcal{M}_b \backslash i)$	0
Bad news	A bandit	All dealers	$\frac{\alpha}{2}Prob(f \in \mathcal{N})$	$-x^s(M_b)Q\frac{(\sigma-S_b)}{2}$
$v_1 = v_0 - \frac{\sigma}{2}$	Dealer i	All other dealers	$\frac{\alpha}{2}Prob(f = i)$	0
	Any dealer k, $k \neq i$	A bandit	$\frac{\alpha}{2}\Phi Prob(f \in \mathcal{M}_b \backslash i)$	$-x^s(M_b - 1)Q\frac{(\sigma-S_b)}{2}$
	Any dealer k, $k \neq i$	All other dealers	$(1-\Phi)\frac{\alpha}{2}Prob(f \in \mathcal{M}_b \backslash i)$	0
No news	A liquidity trader submits a buy order		$\frac{(1-\alpha)\beta}{2}$	$\delta Q\frac{S_b}{2M_b}$
$v_1 = v_0$	A liquidity trader submits a sell order		$\frac{(1-\alpha)\beta}{2}$	$\delta Q\frac{S_b}{2M_b}$
	No order is submitted		$(1-\beta)(1-\alpha)$	0

When there is an innovation in the asset value, good news or bad news, the outcome of the trading game depends on who observes the news first and who is second to observe it. A dealer who observes an innovation immediately updates his quotes. A bandit who observes news immediately submits an order. The table lists all possible outcomes, the probability of each outcome, and dealer i's trading payoff associated with each outcome. When no news occurs a liquidity trader may submit an order. Dealer i's net payoffs are equal to the trading payoff minus the the monitoring cost, which is equal to $\Psi(\lambda_i)$.

Table 2
Bandit j's Payoffs

Innovation in	Outcome of the trading game			
asset value	First trader to observe news	Second trader(s) to observe news	Probability of outcome	Dealer i's trading payoff
Good news	Bandit j	All dealers	$\frac{\alpha}{2} Prob(f = j)$	$Q^s(M_b)\frac{(\sigma - S_b)}{2}$
$v_1 = v_0 + \frac{\sigma}{2}$	Bandit k, $k \neq j$	All Ddealers	$\frac{\alpha}{2} Prob(f \in \mathcal{N} \backslash j)$	0
	A dealer	Bandit j	$\frac{\alpha}{2N} \Phi Prob(f \in \mathcal{M}_b)$	$Q^s(M_b - 1)\frac{(\sigma - S_b)}{2}$
	A dealer	Bandit k, $k \neq j$	$(1 - \Phi)\frac{\alpha}{2} Prob(f \in \mathcal{M}_b)$	0
Bad news	Bandit j	All dealers	$\frac{\alpha}{2} Prob(f = j)$	$Q^s(M_b)\frac{(\sigma - S_b)}{2}$
$v_1 = v_0 - \frac{\sigma}{2}$	Bandit k, $k \neq j$	All dealers	$\frac{\alpha}{2} Prob(f \in \mathcal{N} \backslash j)$	0
	A dealer	Bandit j	$\frac{\alpha}{2N} \Phi Prob(f \in \mathcal{M}_b)$	$Q^s(M_b - 1)\frac{(\sigma - S_b)}{2}$
	A dealer	Bandit k, $k \neq j$	$(1 - \Phi)\frac{\alpha}{2} Prob(f \in \mathcal{M}_b)$	0
No news	A liquidity trader submits a buy order		$\frac{(1-\alpha)\beta}{2}$	0
$v_1 = v_0$	A liquidity trader submits a sell order		$\frac{(1-\alpha)\beta}{2}$	0
	No order is submitted		$(1 - \beta)(1 - \alpha)$	0

When there is an innovation in the asset value, good news or bad news, the outcome of the trading game depends on who observes the news first and who is second to observe it. A dealer who observes an innovation immediately updates his quotes. A bandit who observes news immediately submits an order. The table lists all possible outcomes, the probability of each outcome, and bandit j's trading payoff associated with each outcome. When no news occurs a liquidity trader may submit an order. Bandit j's net payoffs are equal to the trading payoff minus the the monitoring cost, which is equal to $\Psi(\gamma_j)$.

Table 3
Summary Statistics

Variable	Sample 1 (N=310)					Sample 2 (N=200)				
	Mean	Median	Std. Dev.	Min.	Max.	Mean	Median	Std. Dev.	Min.	Max.
Number of SOES clusters	204	61	540	0	6134	15.88	2.00	41.89	0.00	361.00
Frequency of SOES clusters (percent)	1.30	0.93	1.29	0.00	6.23	0.48	0.10	0.89	0.00	5.67
Number of SOES trades	2466	1047	5651	50	62178	222	93	373	2	2598
Number of non-SOES trades	9998	5441	16470	924	151236	2120	1479	1706	703	9871
Bid-ask spread (percent)	1.30	1.14	0.68	0.11	3.97	2.57	2.26	1.32	0.45	6.74
Volatility (percent)	0.87	0.86	0.34	0.16	2.68	0.94	0.90	0.36	0.20	2.65
Maximum SOES quantity	968	1000	140	200	1000	750	750	251	500	1000
Number of dealers	22.28	20.52	10.25	5.33	63.52	13.65	13.02	6.34	3.00	34.95
Average trade size	1666	1483	700	595	5381	1540	1447	592	373	4379
Market capitalization	2457	791	8885	71	107500	472	211	743	12	6651
Average price	26.71	23.53	17.27	5.02	130.94	18.69	14.41	13.79	3.25	71.25

A SOES cluster is defined as an uninterrupted sequence of three SOES trades of maximum size, at the same price, within 30 seconds. The frequency of SOES clusters is measured relative to the total number of trades. The number of SOES trades includes all trades executed in the SOES system irrespective of size. The number of non-SOES trades refers to all trades during regular trading hours that were not submitted to the SOES system. For each stock, the bid-ask spread is measured as the time-weighted average of the relative inside spread. Volatility is measured by the standard deviation of the half-hour returns computed based on the mid-quotes. The maximum SOES quantity is a discrete variable with can take on values 200, 500, or 1000 for sample 1 and 500 and 1000 for sample 2. The number of dealers is computed as a time-series average of the number of active dealer in each stock. The average trade size is measured as the average number of shares per trade excluding any trades that were categorized as a part of a SOES cluster. The market capitalization (millions of dollars) and the average price are based on monthly CRSP data.

Table 4

Correlation Matrix

	soes	spr	vlty	maxq	ndlr	liqd	mkcp
spr	-0.684						
	(-0.507)						
vlty	-0.234	0.487					
	(-0.324)	(0.418)					
maxq	0.281	-0.075	-0.118				
	(0.491)	(-0.130)	(-0.346)				
ndlr	0.108	-0.257	-0.319	0.188			
	(0.003)	(-0.134)	(-0.159)	(0.143)			
liqd	-0.185	0.044	-0.126	-0.733	-0.171		
	(-0.165)	(0.038)	(0.218)	(-0.623)	(-0.145)		
mkcp	0.464	-0.751	-0.524	0.079	0.425	-0.007	
	(0.590)	(-0.710)	(-0.427)	(0.252)	(-0.035)	(-0.044)	
avgp	0.592	-0.757	-0.342	0.057	-0.169	-0.023	0.697
	(0.539)	(-0.625)	(-0.352)	(0.178)	(-0.422)	(-0.126)	(0.790)

The variables in the correlation matrix are the following: the log odds ratio of the probability of a SOES cluster (soes), the average time-weighted bid-ask spread (spr), the maximum SOES quantity (maxq), the number of dealers (ndlr), the average trade size relative to the maximum SOES quantity (liqd), the logarithm of the market capitalization (mkcp), the logarithm of the average price (avgp). The correlation coefficients for the second sample are reported in parentheses directly below the corresponding coefficients for the first sample.

Table 5
Two-Equation Model of SOES Activity and the Bid Ask Spread

	Sample 1 (N=310)				Sample 2 (N=200)			
	SOES Equation		Spread Equation		SOES Equation		Spread Equation	
Variable	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
Constant	-2.52843	0.000	0.05926	0.000	-6.44643	0.000	0.09370	0.000
SOES proxy			0.00436	0.079			0.00085	0.479
Bid-ask spread	-70.85455	0.000			-69.90054	0.000		
Volatility	31.93144	0.000	0.17847	0.058	44.13238	0.072	0.29890	0.115
Maximum SOES quantity	0.00105	0.000			0.00230	0.000		
Number of dealers			-0.00032	0.000			-0.00082	0.000
Liquidity demand			0.00018	0.453			-0.00165	0.003
Market capitalization			0.00023	0.686			-0.00250	0.015
Average price			-0.01097	0.000			-0.01175	0.000

This table reports the parameter estimates with p-values for the two-equation system given in Equation (29) with the logarithm of market capitalization and the average price added as control variables. The system is estimated using three-stage least squares. The dependent variable in the SOES Equation is the SOES proxy, which is defined as the log odds ratio of the probability of a SOES cluster. The explanatory variables in the SOES Equation are: (i) the bid-ask spread, which is defined as the time weighted average of the relative inside spread, (ii) volatility, which is defined as the standard deviation of half-hour mid-quote returns, and (iii) maximum SOES quantity. In the Spread Equation the bid ask spread is the dependent variable and the explanatory variables are: (i) the SOES proxy, (ii) volatility, (iii) number of dealers, which is defined as the time-series average of the number of active dealers, (iv) liquidity demand, which is defined as the average size of trades that are not part of a cluster divided by the maximum SOES quantity, (v) the logarithm of the market capitalization, and (vi) the logarithm of the average price.

Table 6

Reduced Form Equations for SOES Activity and the Bid Ask Spread

	Sample 1 (N=310)				Sample 2 (N=200)			
	SOES Equation		Spread Equation		SOES Equation		Spread Equation	
Variable	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
Constant	-5.19623	0.000	0.03680	0.000	-12.64624	0.000	0.08182	0.000
Volatility	13.36749	0.259	0.24123	0.012	17.54830	0.376	0.30070	0.204
Maximum SOES quantity	0.00116	0.000	0.000004	0.058	0.00267	0.000	0.000004	0.213
Number of dealers	0.02006	0.000	-0.00024	0.000	0.03490	0.011	-0.00085	0.000
Liquidity demand	0.04448	0.138	0.00020	0.333	0.26641	0.001	-0.00095	0.145
Market capitalization	-0.12125	0.006	0.00006	0.857	0.14478	0.187	-0.00244	0.031
Average price	0.77110	0.000	-0.00820	0.000	0.81411	0.000	-0.01095	0.000
R-squared	0.457		0.738		0.526		0.621	
F-test (p-value)	54.36	(0.0000)	99.99	(0.0000)	42.66	(0.0000)	44.59	(0.0000)

This table reports ordinary least-squares parameter estimates (White's robust variance estimator), with corresponding p-values, for the reduced form equations that correspond to the two-equation system given in Equation (29). The dependent variables are the SOES proxy (SOES Equation), defined as the log odds ratio of the probability of a SOES cluster, and the time weighted relative inside bid ask spread (Spread Equation). The explanatory variables are: (i) volatility, which is defined as the standard deviation of half-hour mid-quote returns, (ii) maximum SOES quantity, (iii) number of dealers, which is defined as the time-series average of the number of active dealers, (iv) liquidity demand, which is defined as the average trade size relative to the maximum SOES quantity excluding any trades that were part of a SOES cluster, (v) the logarithm of market capitalization, and (vii) the logarithm of the average stock price.

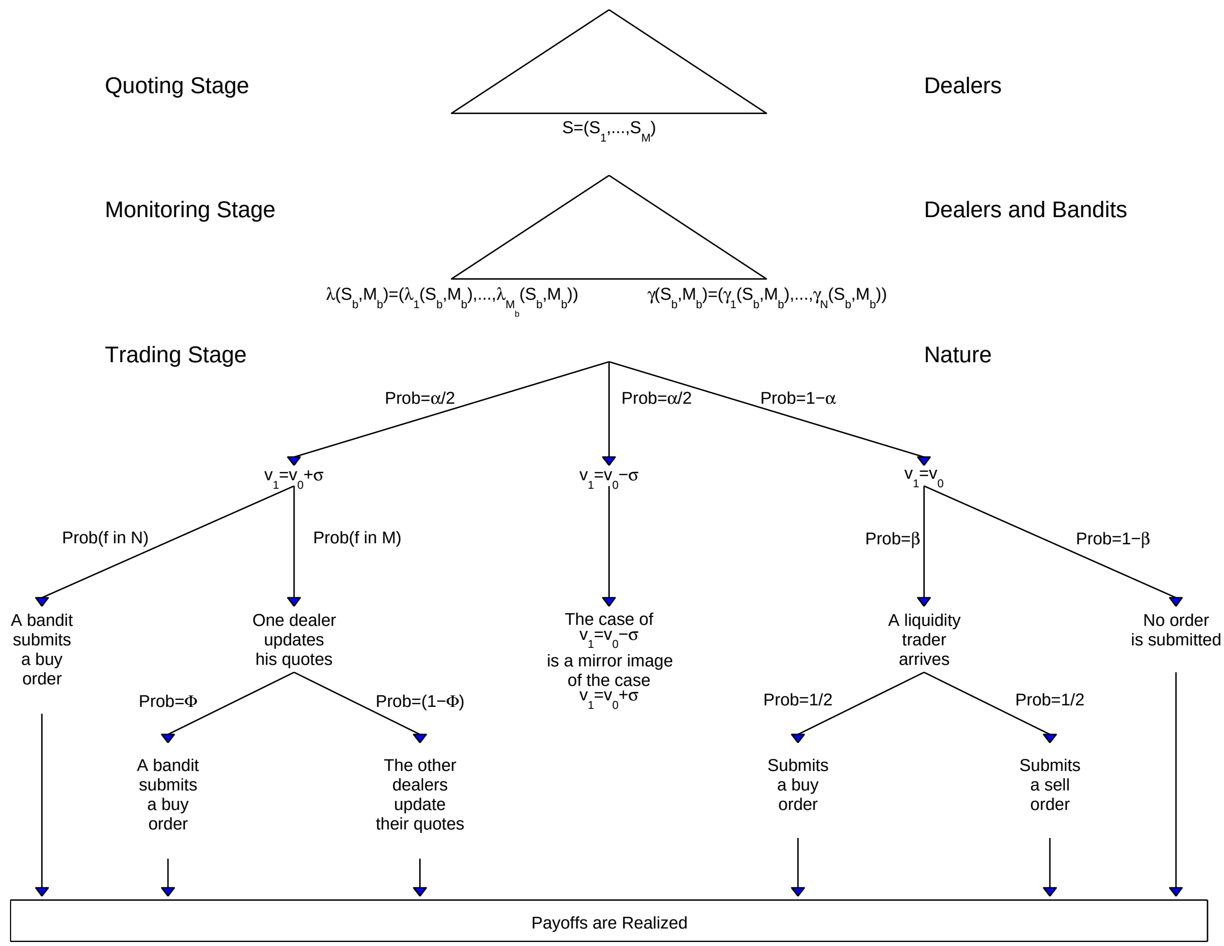

Figure 1: The Trading Game.

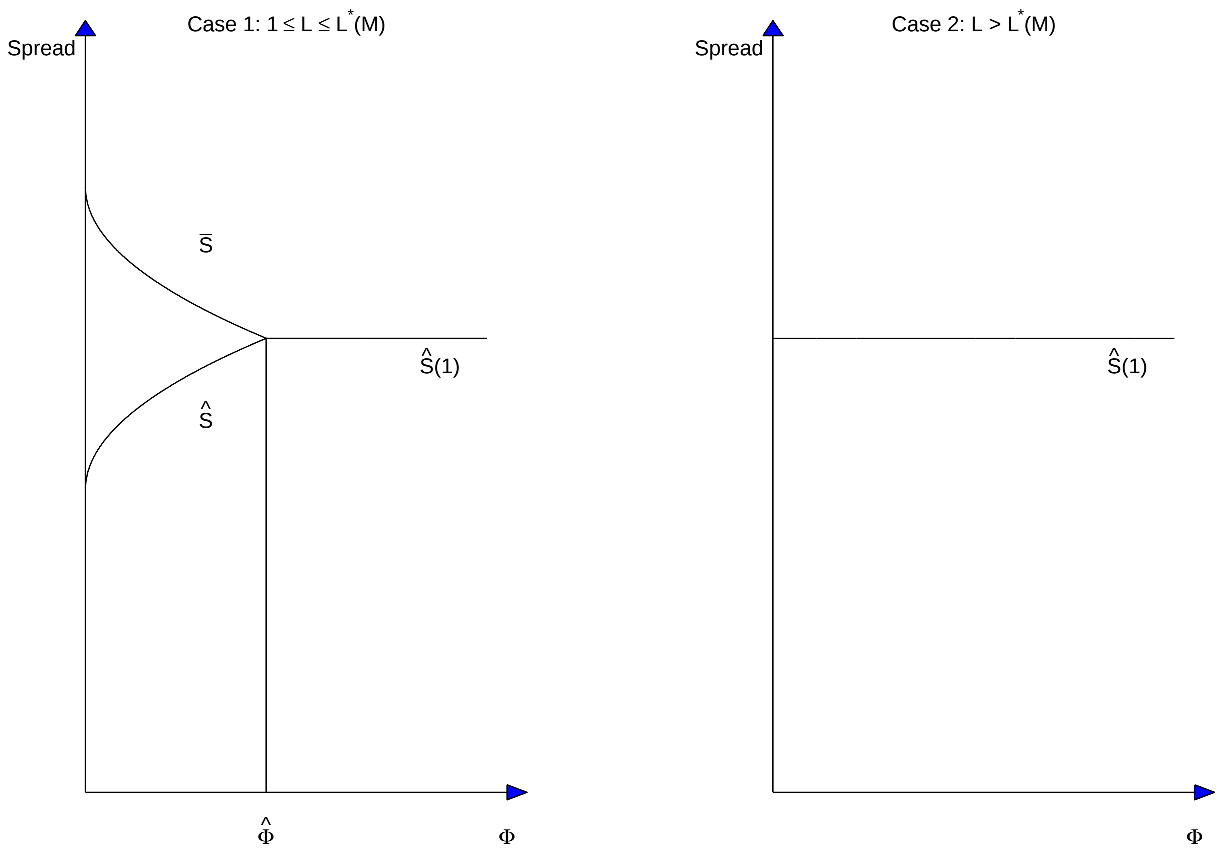

Figure 2: The equilibrium relationship between Φ, L, and the spread.

Legends

Figure 1

In the quoting stage, M dealers quote their spreads $S_1, \ldots, S_M$. The number of dealers quoting the inside spread is denoted by M_b. In the monitoring stage, the M_b dealers who are quoting the inside spread and the N bandits choose their monitoring levels denoted by $\lambda_1, \ldots, \lambda_{M_b}$ and $\gamma_1, \ldots, \gamma_N$, respectively, for the dealers and the bandits. In the trading stage, there is an innovation in the asset value v_0 with probability α. Conditional on a positive innovation, there are three possible outcomes. With probability $Prob(f \in \mathcal{N})$ a bandit submits a buy order. With probability $Prob(f \in \mathcal{M}) \times \Phi$ a dealer updates his quotes and a bandit submit a buy order to the remaining $M_b - 1$ dealers. With probability $Prob(f \in \mathcal{M}) \times (1 - \Phi)$ all dealers update their quotes and no order is submitted by the bandits. The case of a negative innovation is symmetric. With probability $(1 - \alpha)$ there is no innovation and a liquidity buy or sell order is submitted, each with probability $\beta/2$. No order is submitted with probability $(1 - \beta)$.

Figure 2

The spread in zero expected profit equilibrium is denoted by $\hat{S}$. The spread in the maximum profit equilibrium is denoted by $\bar{S}$.

Notes

[1]SOES day traders (bandits) accounted for 83% of SOES share volume as of September 1995, according to the General Accounting Office 1998 report on "The Effect of SOES on the Nasdaq Market."

[2]A Washington Post article (Hinden (1994)) quotes Joseph Hardiman, president of the National Securities Dealers Association, saying that "The SOES activists [SOES bandits] 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." In testimony before the House Committee on Commerce in 1998, David Whitcomb argued that "Abolishing SOES would remove the 'market discipline', which keeps market makers on 'their toes' and causes prices to rapidly adjust when news occurs."

[3]Alternatively, $x^s(M_b)$ and $1/M_b$ can be seen as the probabilities that a dealer receives an order from a bandit or a liquidity trader, respectively.

[4]Houtkin (1998) lists events that SOES bandits monitor: announcements of earnings or economic indicators, price movements in related stocks, and brokerage firms' upgrades and downgrades of stocks.

[5]Results are qualitatively similar when dealers and bandits have different c parameters.

[6]Quote updates are, of course, only *noisy* signals of changes in the value of the asset. However, the logic of the model applies insofar as quote revisions do contain information.

[7]We assume that the inside spread is strictly smaller than the size of the revision in the asset's expected value conditional on information arrival, i.e., $S_b < \sigma$. This is always the case in equilibrium.

[8]Another possibility would be to explicitly model quote revisions. This would make the model much more complex to analyze without adding insights. In any case, the equilibria we describe are robust to the possibility of quote revisions in the sense that no dealer would find it optimal to

unilaterally revise his quotes if he was offered the opportunity to do so (before information arrival, of course).

[9]Second order conditions for the dealers' and bandits' optimization problems are satisfied if $S_b \leq \sigma$, which is the case in equilibrium.

[10]This result is consistent with Harris and Schultz (1998). They find that contrary to the popular view that bandits only pick off the very slow dealers, bandits on average trade before most dealers update their quotes.

[11]Observe also that the bandits and the dealers monitoring levels decrease with the scale of the monitoring cost, c. In equilibrium, the adjustment in monitoring levels exactly offsets the increase in c and the monitoring costs are unchanged. The various picking off probabilities are unaffected as well because the relative monitoring levels do not depend on c. This is why the parameter c does not appear in the bandits' or the dealers' equilibrium expected profits.

[12]Kandel and Marx (1997) show that multiple equilibrium spreads can arise when prices are discrete. Interestingly, we obtain a multiplicity of equilibrium spreads even with continuous prices.

[13]We thank one of the referees for suggesting this interpretation.

[14]To see why, consider a situation in which several dealers post the inside spread and make zero expected profits and assume $L \geq M$ so that $x^s(M) = 1$. If a dealer slightly undercuts, he captures the whole order flow from liquidity traders whereas he keeps trading the same number of shares with bandits (since $x^s(M) = x^s(1) = 1$). Hence the dealer earns a strictly positive expected profit if he undercuts and the situation in which several dealers post the inside spread is not an equilibrium. A similar phenomenon arises in Dennert (1993).

[15]When only one dealer posts the inside spread, the spread does not depend on Φ.

[16]This effect is present in all equilibria of the quoting stage. A firm quote rule also has an indirect effect on the dealers' news monitoring because it affects the equilibrium spread. The direction of the indirect effect depends on the equilibrium in the quoting stage. In the maximal spread equilibrium, the firm quote rule reduces the spread and in this way further increases the

dealers' news monitoring. In contrast, in the zero expected profit equilibrium, the firm quote rule widens the spread and in this way reduces the dealers' need to monitor. Still, this is insufficient for their equilibrium monitoring levels to be smaller than in the case of a relaxed quote rule.

[17]In the model, the probability that one trader will discover an innovation is always equal to one. However, this can be modified so that this probability is less than one, by adding a constant p in the denominators of $Prob(f = i)$ and $Prob(f = j)$. The probability that an innovation will *not* be discovered is then $\frac{p}{\lambda_A+\gamma_A+p}$. It decreases with $(\lambda_A + \gamma_A)$. Thus, the speed of price discovery increases with the aggregate monitoring level.

[18]An integer solution may not exist. In order to avoid this technical problem, we treat N as a real number, as is usual in market entry analysis.

[19]We considered other possible specifications for the number of orders and the interval of time between orders within a cluster. Our empirical results are robust with respect to the different specifications.

[20]The maximum SOES order size is determined by the trading characteristics of the security. Requirements for a 1000 share maximum size include a non-block trading volume of 3000 shares or more per day and three or more market makers. Additional rules require that all IPOs, irrespective of market capitalization and trading volume, trade with a 200 share maximum size for a minimum of 45 trading days. In addition, a security can only move one size category per review.

[21]There are a total of 12 stocks in the first sample and 70 in the second sample for which the total number of clusters is zero. There are fewer zero cluster observations for stocks with the largest SOES size, 8 and 18 for the first and second sample, respectively. To ensure that the log of the odds-ratio is always defined we add one to both the number of clusters and the total number of trades.

[22]The system is estimated using three-stage least squares to account for possible cross-equation correlation in the disturbances and to improve efficiency. Note that if the disturbances are uncorrelated three-stage least squares reduces to two-stage least squares. In our estimations the qualitative effects are unchanged but the coefficient estimates change somewhat suggesting that accounting for

cross-equation correlation is useful. The log of the market capitalization and the average price are added to the spread equation as additional control variables.

[23]Note that due to the non-linear transformation, the exact effect of a change depends on the level of the probability of a SOES cluster.

[24]Harris and Schultz study changes in SOES trading and the average spread around a change in the maximum SOES quantity from 1,000 to 500 shares and find strong evidence of a drop in bandit activity, but little evidence of a drop in the spread.

[25]See for instance Bernhardt, Dvoracek, Hughson, and Werner (2000) for a model of price improvements. Their analysis shows that price improvements are likely to be determined by factors that we can not capture in our analysis (e.g., brokers' identities and brokers' trading frequency with a given dealer).