Journal of Financial Intermediation **8**, 141–173 (1999)
Article ID jfin.1999.0266, available online at http://www.idealibrary.com on IDEAL®

Minimum Price Variations, Time Priority, and Quote Dynamics*

Tito Cordella

IMF and CEPR, 700 19th Street NW, Washington, DC 20431
E-mail: tcordella@imf.org

and

Thierry Foucault†

HEC School of Management and CEPR, 1 rue de la Libération, 78351 Jouy en Josas, France
E-mail: foucault@hec.fr

Received October 29, 1996

We analyze price competition between dealers in a security market where the bidding process is sequential. The model provides an interpretation for the evolution of the best ask and bid prices, in between transactions. We find that convergence to the competitive ask and bid prices can take time. The speed of convergence is determined by the frequency with which dealers check their offers and by the tick size. This creates a relationship between the expected trading cost and the timing of offers posted by the dealers. We also find that a zero minimum price variation never minimizes the expected trading cost. Finally, we study the role of time priority. *Journal of Economic Literature* Classification Numbers: D43, G10. © 1999 by Academic Press

* We are grateful to Dan Bernhardt, Simon Gervais, Maureen O'Hara (the Editor), Christine Parlour, and two referees for their detailed comments, which have helped to improve the paper. Comments from Jeff Bacidore, Bruno Biais, Patrick Bolton, Cécile Boyer, Lawrence Harris, Gerald Garvey, Hugo Hopenhayn, Bryan Routdlege, Chester Spatt, and Jean Tirole have also been useful. In addition, we thank seminar participants from Carnegie–Mellon University, CentER at Tilburg University, the London Business School, HEC School of Management, Toulouse University, the Symposium on Financial Markets at Gerzensee, the 5th Symposium in Market Microstructure at Indiana University, the Western Finance Association Meetings in San Diego, the Conference on High Frequency Data and Market Microstructure in Paris, and Natalie Baumer for editorial assistance. Thierry Foucault acknowledges the financial support of DGICYT Grant PB93-0388. The views expressed here are personal and should not be attributed to the IMF. All errors are ours.

† To whom correspondence should be addressed.

1042-9573/99 $30.00

1. INTRODUCTION

In many financial markets, the quotes are the outcome of a sequential bidding process. As an illustration, consider the following stylized example. Two dealers compete in the market for a security which is worth \$100.10. The minimum price variation (or tick size)[1] in this market is \$1/8. One dealer posts an offer, say an offer to sell, at $\$100\frac{7}{8}$ (i.e., \$100.87). After a while, say 80 s, his or her competitor improves upon this offer, bidding at $\$100\frac{5}{8}$. One minute later, the first dealer realizes that he or she has been undercut. This dealer reacts by posting a new offer to sell at $\$100\frac{1}{2}$. This process proceeds until the arrival of a buy order. This real-world sequential bidding process is *not* adequately captured in the typical market microstructure models. The reason is that these models use Bertrand price competition to formalize this bidding process. This makes the approach static, with dealers assumed to choose their offers simultaneously. Consequently, the existing literature is unable to address the question of how the best offer evolves in between transactions. We address this question by developing a model in which dealers post and update their offer in sequence, as in the previous example. In this way, we relate the timing of offers and the trading costs.

In the model, two dealers post sell prices for one unit of a security. The dealers revise these prices, in turn, at discrete points in time, until the arrival of a market order. A key feature of the model is that the dealers do not know the date at which the market order will be submitted. Consider a dealer who is about to revise his or her offer and who observes that the best price in the market is above the *competitive price* (the expected value of the asset rounded to the nearest tick in the model). The dealer faces the following trade-off. He or she can post the competitive price. In this case, the dealer secures the execution of the next order and a profit equal to the tick size. Or he or she can undercut the current quote by only one tick. Thus, the dealer obtains a larger profit in case of execution, but he or she runs the risk of being undercut by his or her competitor before the arrival of the market order. This risk is small if the competitor does not react too quickly to the new offer. It follows that there are cases in which the dealer is better off undercutting by only one tick. We show that this results in several rounds of sequential price improvements before one dealer posts the competitive price. As a consequence, a transaction can occur at a noncompetitive price and the trading cost depends on the determinants of the speed with which the best price adjusts to the competitive level.

The size of the minimum price variation is one of these determinants. We find that the time for the best price to adjust to the competitive price decreases when the tick size increases. Actually, a larger tick size creates a bigger wedge between the competitive price and the expected asset value, yielding dealers a greater profit.

[1] In security markets, traders must post their quote on a prespecified grid. The minimum price variation is equal to the increment between two consecutive prices on this grid. In most of the cases, the minimum price variation is mandatory and is chosen by market organizers (see Angel, 1997).

Consequently, dealers are more willing to post the competitive price quickly, and convergence to this price is faster. This faster price adjustment means that a larger tick size does not necessarily result in a larger expected trading cost for the liquidity demander. In fact, we establish that the tick size which minimizes the expected trading cost is always strictly greater than zero. Thus, our model provides a new justification for the use of a mandatory tick size in financial markets.

Another important determinant of price dynamics is the frequency with which dealers check their offer. When dealers frequently revise their offer, each dealer realizes that he or she is likely to be quickly undercut if he or she does not post the competitive price. Competitive pressures are stronger and convergence of the best price in the market to the competitive price is faster. An implication is that there exists a negative relationship between the expected trading cost and the frequency of quote changes. We show that this frequency depends on the market-monitoring cost borne by the dealers. Thus ultimately, we establish a connection between the expected trading cost for liquidity demanders and the market-monitoring cost incurred by liquidity suppliers. We also obtain the result that the optimal minimum price variation increases with the monitoring cost.

In dynamic trading environments, time priority is often used to allocate a trade between liquidity suppliers posting the best price.[2] By this rule, at a given price, orders entered first are executed first. We compare the equilibria obtained when time priority is enforced with those obtained with another allocation rule: if there is a tie, the order is split equally among the dealers posting the best price. In this case, we find that dealers can stop undercutting each other at a price way above the competitive price. Bidders are deterred from improving upon their competitor's offer by the mere threat of triggering a price war, which would ultimately depress the trading profits of all the dealers. In contrast, when time priority is enforced, dealers keep undercutting each other until one of them posts the competitive price. Thus, after a while, the competitive price is *eventually* posted with time priority.

Maskin and Tirole (1988) also consider a model in which agents alternate in posting prices. They show that noncompetitive prices can be obtained. They do not analyze the role of the tick size and the frequency with which prices are revised, variables which play a central role in our model. Furthermore, since we use a different set of assumptions concerning the market structure (e.g., time priority), our results differ substantially from those of Maskin and Tirole (1988). For example, in our setting, the best price in the market converges to the competitive price, which is never the case in their analysis. Dutta and Madhavan (1997) show that collusive outcomes can be sustained when dealers compete over time. They consider a repeated game in which dealers post prices *simultaneously*, at different points in time. In our framework, the dealers post offers *in sequence*. This provides an important role for time priority, and we show that this priority rule makes

[2] Domowitz (1993) shows that this rule is prevailing in many electronic markets. There are important exceptions, however. Time priority is generally not enforced in open-outcry markets like the CBOT. It is not enforced on the NASDAQ either.

(implicit) collusive pricing more difficult. Finally, to our knowledge, this model is the first to analyze the effect of monitoring costs on the timing of offers in a model of dynamic price competition.

Anshuman and Kalay (1998), Chordia and Subrahmanyam (1995), and Kandel and Marx (1997) study static models of price competition in which the minimum price variation plays an important role. They show that price discreteness allows dealers to capture some rents. They all conclude that a *zero* minimum price variation eliminates dealers' rents, i.e., minimizes the expected trading cost for liquidity demanders. Bernhardt and Hughson (1996) also obtain this result in a model in which the dealers post their offers sequentially. In all these models, a decrease in the minimum price variation reduces the wedge between posted prices and dealers' reservation prices. For this reason, such a decrease lessens the expected trading cost. Although this effect is present in our framework, it is counterbalanced by another effect: a decrease in the tick size lengthens the time it takes for the best price to reach the competitive level. Since a market order can arrive in between the dates at which dealers revise their offers, it follows that a zero tick size never minimizes the expected trading cost.

The next section spells out the model and the equilibrium concept which are used to solve the trading game. Section 3 characterizes dealers' bidding strategies. The determinants of (i) price dynamics and (ii) the expected trading cost are also analyzed. Section 4 endogenizes the frequency with which dealers revise their offers. Section 5 analyzes the role of time priority. Section 6 discusses the robustness of the results. Section 7 concludes.

2. THE MODEL

In this section, we describe a model in which price competition between dealers is dynamic, and we present the equilibrium concept that we use to solve the trading game.

2.1. The Traders and the Trading Process

Consider the market for a security. Time is continuous and is indexed by $t \in [0, +\infty)$. Prices are posted by two risk-neutral dealers. For brevity, we focus on the equilibrium ask prices. It is straightforward to extend the results to the case in which dealers post ask *and* bid prices.

Dealers post an offer and revise it sequentially at dates $\{0, 1, \ldots, \tau, \tau+1, \ldots\}$. For instance, dealer 1 (say) posts a new ask price a^1_τ at date τ. Then, after a period of time $\Delta > 0$, dealer 2 reacts, posting a new offer $a^2_{\tau+1}$. This alternating bidding process is repeated until a trade occurs. The dealer who posts the first offer at date 0 is randomly chosen. The period Δ is the time it takes for a dealer to react to a new offer posted in the market. We refer to Δ as the *reaction time*. The smaller the reaction time, the more frequent are the quote revisions by the dealers. Checking offers takes time, and for this reason dealers refrain from revising their quotes too frequently. In Section 4, we endogenize the reaction time using this idea. To simplify the exposition, we have assumed that Δ is deterministic. It is worth pointing out, however, that all the results that follow can be derived assuming that the reaction time is random. In this case, the average reaction time plays the role of Δ.

A transaction occurs when a trader wants to buy the asset for liquidity reasons. For simplicity, we assume that the liquidity buyer has a rectangular demand.[3] Namely, he or she purchases L units of the asset if the best ask price is lower than or equal to R_B. Thus, R_B is the maximum ask price that can be quoted by the dealers. We normalize L to 1. Furthermore, the date $\tilde{T}$ at which a liquidity trader submits his or her buy market order is *random*. We assume that $\tilde{T}$ is exponentially distributed with parameter λ.[4] Consequently, in each interval $[\tau, \tau+1]$, there is a probability $\Phi(\Delta) = 1 - e^{-\lambda\Delta}$ that a buy market order arrives.

When the trade occurs, the game stops, and the asset payoff, $\tilde{V}$, is realized. The uncertainty concerning the payoff date guarantees that the model is stationary, which simplifies the derivation of the equilibrium bidding strategies. The expected value of the asset is denoted μ. Figure 1 provides a graphical representation of the timing of the trading process.

2.2. The Trading Rules

Minimum price variation. The set of possible quotes is discrete. The grid on which dealers can post their prices is characterized by the size of the minimum price variation, $g(n) = \frac{g}{2n+1}$, $n \in \mathbb{N}$. Notice that the larger n is the finer the grid is. By varying n, we are able to compare equilibria obtained with different grids. A grid with a zero minimum price variation is obtained as a limiting case by taking n to infinity. From now on, when we say that the tick size is decreased, it must be understood that n is increased. On the grid with tick size $g(n)$, we denote by $\langle p \rangle_n^-$ the highest price which is strictly lower than p. In a similar way, on the grid with size $g(n)$, $\langle p \rangle_n^+$ is the lowest price which is greater than or equal to p. The set of possible prices with a grid of size $g(n)$ is $\mathcal{P}_n = \{\ldots, p(-i), \ldots, p(0), \ldots, p(i), \ldots\}$, with $p(i) = \langle \mu \rangle_n^- + ig(n)$ and $p(-i) = \langle \mu \rangle_n^- - ig(n)$, $i \in \mathbb{N}$.

We assume that $\mu - \langle \mu \rangle_0^- = \langle \mu \rangle_0^+ - \mu = \frac{g(0)}{2}$; that is, the position of the asset expected value is halfway between ticks for the grid with size $g(0)$. We also assume that the liquidity trader's reservation price is on this grid; that is, there exists $k \in \mathbb{N}$ such that $R_B = \langle \mu \rangle_0^- + kg(0)$, with $k \geq 2$ for the problem to be

[3] In a previous version of the paper, we analyzed the case in which the order flow was price-dependent. The results are qualitatively the same as those obtained in this simpler framework.

[4] Notice that it means that the number of buy market order arrivals in a given interval of time follows a Poisson process with rate λ. This is a standard representation of the order arrival process in the market microstructure literature. This assumption is used, for instance, in Garman (1976) and in Easley *et al.* (1996). Easley *et al.* (1996), and Gouriéroux *et al.* (1997) provide estimate of the arrival rates.

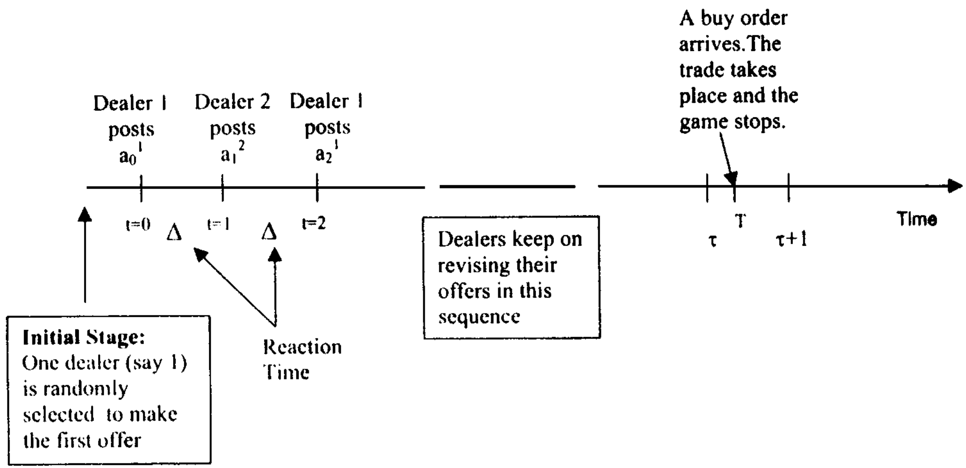

FIG. 1. The trading game. This figure presents the timing of the trading game described in Section 2.

of interest. It is straightforward to check that these two assumptions and the definition of $g(n)$ imply that the asset expected value is always halfway between ticks for all the possible grids and that the liquidity trader's reservation price belongs to $\mathcal{P}_n$, $\forall n \in \mathbb{N}$. These assumptions do not affect the results qualitatively, but they simplify the comparative statics between equilibria obtained with different grid sizes.

Priority rules. Price priority is enforced; that is, a buy market order is executed against the best quote (price priority rule). If there is a tie at the best price, time priority is used. The dealer who is the first to post the price at which the tie occurs executes the market order.

Timing of order submission. An important feature of the model is that the arrival date of the buy market order is random. This is a natural assumption when dealers must post their quotes before observing the arrival of a market order. This timing is the hallmark of quote-driven markets (see Bernhardt and Hughson (1996) for a discussion) such as the NASDAQ, where dealers post quotes valid up to a maximum quantity before being contacted by brokers. This is also the timing of orders in continuous limit order markets (e.g., the Paris Bourse or the Stockholm Stock Exchange).

Order display. We assume that at any point in time, only the best price in the market is publicly displayed. This implies that a dealer who has strict priority of execution does not observe the quote posted by the other dealer.[5] As explained below, in our model, this trading rule restricts the set of possible equilibria: it prevents dealers from using bidding strategies that sustain very noncompetitive outcomes.

2.3. A Benchmark: The Competitive Price

We call $p(1)$ the *competitive price*. This is the first price on the grid above dealers' reservation price (which is the asset expected value in our setting). The competitive price is such that it cannot be undercut profitably.

2.4. Pricing Strategies and Equilibrium Definition

In this subsection, we define the equilibrium concept which is used in this article. This definition is not straightforward, because the game is dynamic and only the best quote in the market is displayed.

2.4.1. Dealers' Objective Functions

Let a_t^j be the quote of dealer j at date t. The best offer in the market at date t is denoted $a_t^m \equiv \text{Min}\{a_t^1, a_t^2\}$. Because dealers alternate in quoting prices, we have $a_t^j = a_\tau^j$, $\forall t \in [\tau, \tau + 2)$, if τ is the last date at which dealer j had the opportunity

[5] This trading rule is used in some markets. For instance, the limit order books on the New York Stock Exchange and the Tokyo Stock Exchange are closed: only the price and size of the best quotes are displayed in these markets.

to make an offer. By convention, index τ is used for an offer posted in the market *after* the reaction of the dealer revising his or her quote at date τ.

Let $\Pi_j(a_t^h, q_t^j, a_t^j)$ be dealer j's expected profit, conditional on the *arrival of a buy market order at time t*. Dealer j's expected profit upon execution depends on the position of his or her quote (a_t^j) relative to his or her competitor's quote (a_t^h), and if there is a tie, on his or her priority status. This priority status is tracked by the indicator variable q_t^j, which takes the value 0 when dealer j does not have time priority and which is equal to 1 otherwise. For all $h, j \in \{1, 2\}, h \neq j$, we get

$$\Pi_j\left(a_t^h, q_t^j, a_t^j\right) = \begin{cases} 0, & \text{if } a_t^j > a_t^h; \\ 0, & \text{if } a_t^j = a_t^h, \text{ and } q_t^j = 0; \\ \left(a_t^j - \mu\right), & \text{if } a_t^j = a_t^h, \text{ and } q_t^j = 1; \\ \left(a_t^j - \mu\right), & \text{if } a_t^j < a_t^h. \end{cases} \tag{1}$$

We call $s_{\tau-1} \equiv \{a_{\tau-1}^m, q_{\tau-1}\}$ the *state of the market* at date τ. The state of the market at date τ is (i) the best quote in the market just before the quote revision at date τ and (ii) the priority status of the dealer who revises his or her quote at this date.[6] For completeness, we set $s_{-1} = \{R_B, 1\}$. The trading history at date τ, $H_{\tau-1}$, is the set of states of the market until this date. Formally, $H_{\tau-1} = \bigcup_{k=-1}^{k=\tau-1} s_k$.

A bidding strategy specifies the quote chosen by a dealer, each time the dealer revises his or her offer. For given bidding strategies of the two dealers, let $E(\Pi_j(a_{\tilde{T}}^h, q_{\tilde{T}}^j, a_{\tilde{T}}^j) \mid \tilde{T} \geq \tau\Delta, H_{\tau-1})$ be dealer j's expected profit at date τ, conditional on the *trading history*. We obtain

$$\begin{aligned} &E\left(\Pi_j\left(a_{\tilde{T}}^h, q_{\tilde{T}}^j, a_{\tilde{T}}^j\right) \mid \tilde{T} \geq \tau\Delta, H_{\tau-1}\right) \\ &\quad = \sum_{k=\tau}^{+\infty} \Phi(1-\Phi)^{(k-\tau)} \Pi_j\left(a_k^h, q_k^j, a_k^j\right) \quad \forall h, j \in \{1, 2\}, h \neq j. \end{aligned} \tag{2}$$

This is dealer j's objective function at date τ.

2.4.2. *Dealers' Bidding Strategies*

Consider the dealer who is about to revise his or her offer at date τ. Only the best price in the market is displayed. Thus the dealer does not necessarily observe the offer posted by his or her competitor. Let $\hat{a}(H_{\tau-1})$ be the belief of this dealer regarding the current offer of his or her competitor. We call $\{s_{\tau-1}, \hat{a}(H_{\tau-1})\}$ the *state of the trading process* at date τ. A bidding strategy is a mapping $R(.)$ from the set of possible states of the trading process into the set of possible offers $\mathcal{P}_n$. The reaction function $R(.)$ gives the new price posted by a dealer when he or she has the opportunity to revise his or her offer, given the state of the trading process.

[6] This means that if dealer j revises his or her quote at date τ then $q_{\tau-1} = q_{\tau-1}^j$.

Note that we focus our attention on Markov strategies, that is, bidding strategies which are functions of the state of the trading process but not directly functions of the trading history. Moreover, we only consider symmetric strategies: for a given state of the market, dealers 1 and 2 revise their offers in the same way.

We are interested in the Markov equilibria of the trading game. The dealers' bidding strategies form a *Markov equilibrium* if (i) the strategies are Markov, (ii) the dealers form their belief on their competitor's offer as explained below, and (iii) for each possible state of the trading process, the offer prescribed by his or her bidding strategy maximizes a dealer's expected profit, given the subsequent actions of the dealer and his or her rival. Thus, a Markov equilibrium is a subgame perfect equilibrium in which traders use Markov strategies.

2.4.3. Dealers' Beliefs

Consider $R(.)$, an equilibrium reaction function. Dealer j is about to revise his or her offer at date τ. Two situations are possible. Either dealer j's competitor is posting the best price in the market or dealer j is posting this price. In the first situation, dealer j observes his or her competitor's offer. Consequently $\hat{a}(H_{\tau-1}) = a^m_{\tau-1}$. In the second situation, he or she does not observe the competitor's offer. However, if $H_{\tau-1}$ is on the equilibrium path, then dealer j's belief must be consistent with the equilibrium reaction of his or her competitor at the previous date; i.e., $\hat{a}(H_{\tau-1}) = R(s_{\tau-2}, \hat{a}(H_{\tau-2}))$. The next lemma establishes that in this case, $\hat{a}(H_{\tau-1})$ is just a function of p, the current best price in the market.

LEMMA 1. *Consider a dealer who is about to revise his or her offer at date τ and who has strict priority at the best quote p. In this case, if the trading history $H_{\tau-1}$ is on the equilibrium path then $\hat{a}(H_{\tau-1}) = R(p, 0, p)$.*

If $H_{\tau-1}$ is out of the equilibrium path and dealer j does not observe his or her competitor's offer, then his or her belief can be chosen arbitrarily. We assume that $\hat{a}(H_{\tau-1}) = a^m_{\tau-1} + g(n)$ in this case. This choice is without consequence for our results (see discussion on uniqueness in Section 3). Thus, a dealer's belief about his or her competitor's offer depends only on the best price in the market. This means that the reaction function can be written as a function of the state of the market. This allows us to simplify our notation. From now on, let $R(p, 0)$ be the optimal reaction of a dealer when he or she observes the state of the market $s_{\tau-1} = \{p, 0\}$, and let $R(p, 1)$ be the dealer's optimal reaction when he or she observes the state $s_{\tau-1} = \{p, 1\}$. It is implicit that if $\{p, 1\}$ is on the equilibrium path, the dealer believes $R(p, 0)$ to be his or her competitor's quote (Lemma 1), while if $\{p, 1\}$ is not on the equilibrium path, the dealer believes $p + g(n)$ to be his or her competitor's quote.

2.4.4. Markov Equilibrium: A Formal Definition

Given a reaction function $R(.)$, let $V(s_{\tau-1})$ be a dealer's expected profit given that (i) the state of the market is $s_{\tau-1}$; (ii) the dealer is about to react; and

(iii) from date $\tau+1$ on, the two dealers will behave according to the reaction function $R(.)$. We also define $W(s_{\tau-1}, a_{\tau-1})$, a dealer's expected profit at date τ in state $s_{\tau-1}$, provided that the dealer chose $a_{\tau-1}$ at the previous date and that his or her *competitor* is about to react. Suppose that dealer j is about to revise his or her offer at date τ. $V(.)$ can be expressed by the dynamic programming relationship

$$V(s_{\tau-1}) = \max_{a_j \in \mathcal{P}_n} \Phi \Pi_j \left(\hat{a}_\tau, q_\tau^j, a_j\right) + (1-\Phi) W(s_\tau, a_j), \tag{3}$$

and $W(., .)$ is

$$W(s_\tau, a_j) = \Phi \Pi_j \left(R(s_\tau), q_{\tau+1}^j, a_j\right) + (1-\Phi) V(s_{\tau+1}), \tag{4}$$

where the evolution of the state of the market between dates τ and $\tau+1$ is determined by dealer j's offer at date τ and henceforth by the actions prescribed by the reaction functions. A reaction function $R(.)$ is a Markov equilibrium iff $R(s_{\tau-1})$ is the solution of Eq. (3) for each state $s_{\tau-1}$, given that dealers' beliefs on their competitor's quote are specified as explained in Subsection 2.4.3.

Following Maskin and Tirole (1988), we call a *focal price* a price p on the equilibrium path such that $R(p, 1) = p$ and $R(p, 0) = p$. If there exists a focal price, once it is reached, the dealers keep posting this price until the arrival of the market order. Finally, we denote by $V(\Delta, g(n))$ a dealer's expected trading profit at the beginning of the trading process. When each dealer has an equal probability of being the first dealer to post an offer at date 0, we obtain

$$V(\Delta, g(n)) = \frac{1}{2} V(\{R_B, 1\}) + \frac{1}{2} V(s_0^*), \tag{5}$$

where s_0^* is the state of the market that is observed, in equilibrium, by the dealer who posts an offer at time 1.

3. QUOTE DYNAMICS AND TRADING COSTS IN EQUILIBRIUM

In this section, we derive the equilibrium bidding strategy. Then we analyze the determinants of the evolution of the best quote and the expected trading cost for the liquidity demander.

3.1. Bidding Strategies in Equilibrium

Consider a date in which dealer j revises his or her offer. The best price in the market is strictly above $p(1)$. Also assume that dealer j does not have execution priority at this price. Dealer j can choose to improve the best price by one or several ticks, thereby capturing price and time priority. Other possibilities for dealer j are to match the best price or to post a higher price than the best price. In these two cases, he or she surrenders price and time priority for the next period. The next proposition establishes that improving upon the current quote is always optimal for dealer j.

PROPOSITION 1. *Consider an equilibrium reaction function $R(.)$. For all values of the parameters, this reaction function has the following properties:*

- **(P1)** $R(p,0) < p$ *if* $p \geq p(2)$; *and*
- **(P2)** $R(p(1),1) = p(1)$, *and* $R(p(1),0) = p(1)$.

Proposition 1 is important for two reasons. First, (P1) states that in equilibrium, the best quote must decrease as long as it is greater than the competitive price. This is the *unique possible evolution* for the best quote. Second, (P2) claims that, with time priority, the *unique focal price* is the competitive price. These results imply that, for each grid, there necessarily exists a price $p \in (p(1), R_B]$, such that when the best quote reaches p, the dealer without execution priority finds it optimal to post $p(1)$ and to hold on to this offer until the arrival of the market order. The next proposition characterizes the price at which the "jump" to the competitive price occurs. It also provides dealers' bidding strategy in equilibrium. We denote the greatest integer strictly lower than x by $\lfloor\lfloor x \rfloor\rfloor$.

PROPOSITION 2 (Equilibrium Bidding Strategy). *The following reaction function, with $p^* = \langle \mu + \frac{g}{2\Phi} \rangle_n^+ = p(1) + \lfloor\lfloor \frac{(\Phi+1)}{2\Phi} \rfloor\rfloor g(n)$, is a Markov equilibrium when time priority is enforced.*

1. $R(p,0) = p - g(n)$, *if* $p^* + g(n) \leq p \leq R_B$.
2. $R(p,0) = p(1)$, *if* $p(1) \leq p \leq p^*$.
3. $R(p,1) = p$, *if* $p^* \leq p \leq R_B$.
4. $R(p,1) = p(1)$, *if* $p(1) \leq p \leq p^* - g(n)$.

The intuition for Proposition 2 is as follows. Consider a dealer who is about to revise his or her offer. The best price in the market is $p + g(n)$. Suppose that the dealer conjectures that if he or she does not quote $p(1)$ in this round, his or her competitor will post the competitive price in the next round. The dealer faces the following trade-off. If the dealer quotes the competitive price, he or she secures execution and obtains with certainty a profit equal to $p(1) - \mu = g(n)/2$. If the dealer undercuts by only one tick, he or she obtains a larger profit, $(p-\mu)$, in case of execution. But the dealer runs the risk of being undercut by his or her competitor. The probability that a market order arrives, before his or her competitor reacts, is Φ. Thus the dealer's expected profit is $\Phi(p-\mu)$, if he or she undercuts by only one tick. It follows that the dealer must undercut by only one tick if

$$\Phi(p-\mu) \geq \frac{g(n)}{2}. \tag{6}$$

The left-hand side of this inequality increases with p. The smallest price on the grid such that the inequality is satisfied is p^* (as defined in Proposition 2). This explains why, if $p \geq p^* + g(n)$, no dealer finds it optimal to undercut the best price in the market by more than one tick. This also implies that the price at which the jump to the competitive price occurs cannot be lower than p^*.

NUMERICAL EXAMPLE. Assume $g = 1$, $\mu = 99.5$, $R_B = 103$. Furthermore, we choose the unit of time to be equal to the expected waiting time between market order arrivals so that $\lambda = 1$. Finally, let us assume that dealer 1 makes the first offer. We will use this parameterization in all the numerical examples. If the grid size is $g(0) = 1$ and $\Delta = 0.5$ (i.e., the reaction time is half the expected waiting time between market orders arrivals), then $p^* = 101$. Dealer 1 first offers a price equal to 103. His or her competitor reacts with a price equal to 102. Then dealer 1 revises his or her initial offer, asking 101. Eventually dealer 2 posts the competitive price equal to 100. The jump to the competitive price takes place at date $\tau = 3$, which means that 1.5 units of time are necessary to reach the competitive level. If, for instance, the market order arrives in between dates 1 and 2, the price paid by the liquidity trader is 102. This shows how, at equilibrium, noncompetitive transaction prices can be obtained when dealers bid sequentially.

Uniqueness. The proof of Proposition 1 does not rely on the specification of dealers' beliefs out of the equilibrium path. Furthermore, the price p^* at which the jump to the competitive price occurs is unique and independent from dealers' beliefs out of the equilibrium path. Thus, in equilibrium and for given values of the parameters, there is a unique possible path for the best ask price. It might look surprising that there are no equilibria in which the best price in the market is raised. Actually, consider the following potential objection: "Assume that the best price has reached the competitive level and that dealer 1 does not have time priority. Dealer 1 can raise his or her quote at, say, $p' > p(2)$. Upon observing this offer, dealer 2 reacts by posting $p' - g(n)$. Both dealers have an incentive to act in this way since competition restarts from a high price level. Thus there should be equilibria in which dealers find it optimal to raise their offers." This argument does not hold for the following reason. It relies on the possibility for dealer 2 to *observe* the level at which dealer 1 has raised his or her price. When this is not the case, dealer 1 has an incentive to deviate and to post $p' - g(n)$ in the first place. This deviation is always optimal since (i) it cannot alter dealer 2's bidding strategy (dealer 2 does not observe the deviation) and (ii) it allows dealer 1 to capture time priority when dealer 2 raises his or her price at $p' - g(n)$. The deviation precludes equilibria in which dealers raise their offers. Since we have assumed that only the best price in the market is displayed, these equilibria do not exist in our model.[7]

[7] In contrast, in the model of dynamic price competition analyzed by Maskin and Tirole (1988), there are equilibria in which firms raise their prices when the market price reaches a sufficiently low level. These equilibria are called "Edgeworth cycles" by Maskin and Tirole. As we just explained, we do not obtain these cycles because, in contrast with Maskin and Tirole (1988), we assume that agents only observe the best price posted in the market.

3.2. Dynamics of the Best Quote

Proposition 1 establishes that the best quote in the market converges to the competitive price. This convergence may take time, however. Consider again Proposition 2. The number of offers and counteroffers that are needed to reach the competitive level depends on the position of p^* relative to R_B. If $p^* \leq R_B$, the dealers undercut each other by only one tick as long as the best price in the market is larger than p^*. Thus, the date τ^* at which the best price becomes p^* is such that $R_B - \tau^* g(n) = p^*$. It follows that

$$\tau^* = \frac{R_B - p^*}{g(n)}. \tag{7}$$

The jump to the competitive price occurs at date $\tau^* + 1$. Alternatively if $p^* > R_B$, the dealer who makes the first offer posts the competitive price and convergence is immediate. This results in the following observations. When the spread is large, small quote improvements are observed in succession. When the spread is small, a large quote improvement is observed and no subsequent improvement takes place. This qualitative description of the evolution of the best price, given the size of the inside spread, is consistent with the empirical findings in Biais *et al.* (1995).

Let $T^*(\Delta, g(n))$ be the time it takes for the best ask price to hit the competitive level. We refer to T^* as the *hitting time*,

$$T^*(\Delta, g(n)) = \mathrm{Max}\{\tau^* + 1, 0\}\Delta, \tag{8}$$

which gives

$$T^*(\Delta, g(n)) = \mathrm{Max}\left\{\frac{R_B - \mu}{g(n)} + \frac{1}{2} - \left\lfloor\left\lfloor \frac{\Phi + 1}{2\Phi} \right\rfloor\right\rfloor, 0\right\}\Delta. \tag{9}$$

We study below the impact of the tick size and the reaction time on the hitting time. As shown in Section 3.3, this proves to be important for understanding the impact of these variables on the trading cost.

Since Φ increases with the reaction time, it is straightforward that the hitting time also (weakly) increases with the reaction time. Let $\overline{\Delta}(g(n))$ be the largest value of the reaction time such that $T^* = 0$. $\overline{\Delta}(g(n))$ is the solution of the following equation:

$$\frac{R_B - \mu}{g(n)} + \frac{1}{2} - \left\lfloor\left\lfloor \frac{\Phi(\overline{\Delta}) + 1}{2\Phi(\overline{\Delta})} \right\rfloor\right\rfloor = 0. \tag{10}$$

It follows that $\overline{\Delta}(g(n)) = -\frac{1}{\lambda}\ln(1 - \frac{g(n)}{2}(R_B - \mu)^{-1})$. For all the pairs $(g(n), \Delta)$ such that $\Delta \leq \overline{\Delta}(g(n))$, the hitting time is equal to zero; that is, the first dealer who makes an offer posts the competitive price. For all the pairs $(g(n), \Delta)$ such that $\Delta > \overline{\Delta}(g(n))$, the hitting time is strictly positive. Figure 2 represents graphically the sets of values for the tick size and the reaction time such that the hitting time is zero

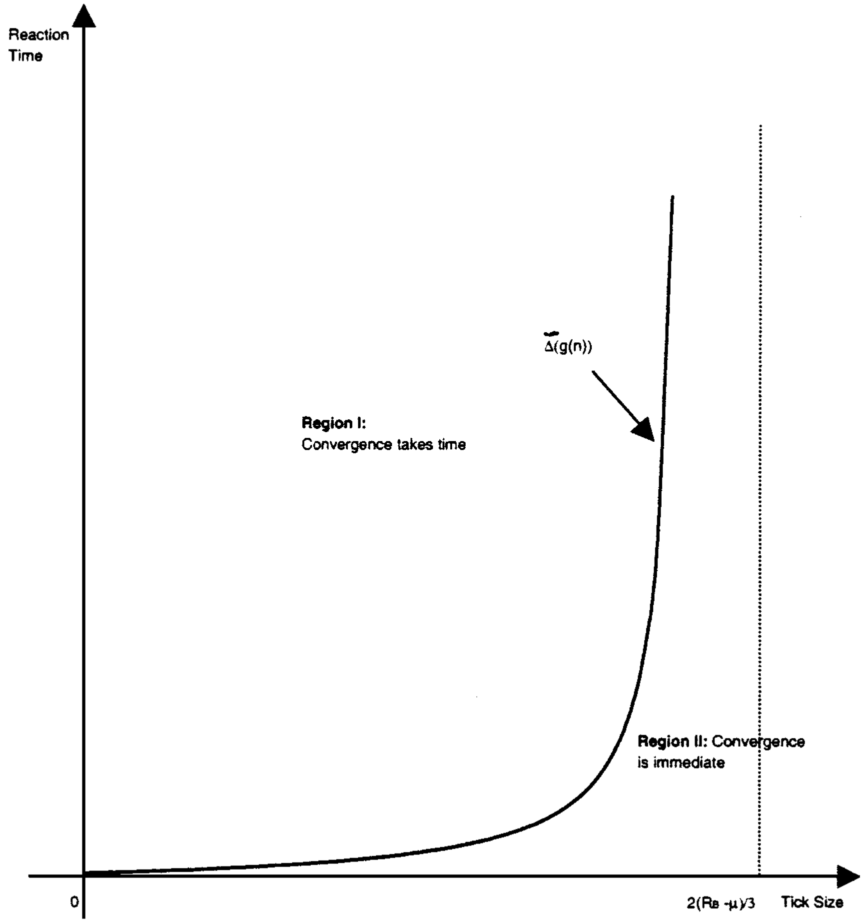

FIG. 2. Tick size, reaction time, and speed of convergence. This figure indicates for each pair $(\Delta, g(n))$ whether convergence to the competitive price is immediate or whether it takes time.

or strictly positive. Note that, for a given reaction time, a decrease in the tick size ultimately results in a strictly positive hitting time. Furthermore, $\bar{\Delta}(0) = 0$, which means that when the tick size is zero, the hitting time is always strictly positive. It is straightforward to show that the hitting time (weakly) decreases with the tick size and that $\lim_{g(n)\to 0} T^*(\Delta, g(n)) = +\infty$, $\forall \Delta > 0$. The next corollary summarizes these results. The discussion, which follows, provides the economic intuition.

COROLLARY 1 (Hitting Time). *In equilibrium,*

1. *The dealer who makes the first offer at date $\tau = 0$ posts the competitive price if and only if the reaction time $\Delta \leq \bar{\Delta}(g(n))$ $\forall g(n) \geq 0$.*
2. *The hitting time (weakly) increases with dealers' reaction time.*

3. *The hitting time (weakly) decreases with the tick size.*
4. $\lim_{g(n)\to 0} T^*(\Delta, g(n)) = +\infty, \forall \Delta > 0.$

A decrease in dealers' reaction time increases the chance that an offer above the competitive level will be undercut before the arrival of a market order. Thus, smaller reaction times induce dealers to more rapidly post the competitive price. To illustrate this point, consider the numerical example of the previous subsection. Dealer 1 is about to revise his or her offer and the best price is 102. Dealer 2 has priority of execution at this price. Dealer 1 must choose between posting a price equal to 101 or posting the competitive price equal to 100. If he or she posts a price equal to 100, dealer 1 secures a profit equal to $p(1) - \mu = \frac{g(0)}{2} = 0.5$. If he or she posts a price equal to 101, dealer 1 realizes a profit equal to $101 - 99.5 = 1.5$, in case of execution. Execution is uncertain, however, since dealer 1 might be undercut by dealer 2. The optimal bidding strategy depends on dealers' reaction time. For instance, if $\Delta = 0.5$, there is a probability $\Phi(0.5) = 0.39$ that the market order will arrive before the reaction of dealer 2. Thus, if he or she posts a price equal to 101, dealer 1 obtains an expected profit equal to $0.39(101 - 99.5) = 0.585 > 0.5$. Dealer 1 chooses an offer equal to 101 in this case. If $\Delta = 0.3$, the probability that the market order arrives before the reaction of dealer 2 is smaller ($\Phi(0.3) = 0.25$). It follows that dealer 1's expected profit with an offer at 101 becomes $0.25(101 - 99.5) = 0.375 < 0.5$. Dealer 1 is now better off posting the competitive price. Consequently, the jump to the competitive price takes place earlier when $\Delta = 0.3$ than when $\Delta = 0.5$. Accordingly, the hitting time is lower when $\Delta = 0.3$ than when $\Delta = 0.5$. Computations show that $p^* = 102$ when $\Delta = 0.3$. This means that $\tau^* = 1$ in this case instead of $\tau^* = 2$ when $\Delta = 0.5$. Thus, $T^*(0.3, 1) = 0.6$, whereas $T^*(0.5, 1) = 1.5$.

A decrease in the tick size translates into a lower profit at the competitive price and reduces the attractiveness of this offer. Thus, dealers are less tempted to quickly post the competitive price in order to lock in a sure profit, equal here to half the tick size. Consider again the numerical example that we discussed in the previous paragraph. Assume that $\Delta = 0.3$ and that the tick size is now $g(1) = 1/3$. The best price in the market is $101 + \frac{1}{3}$. If dealer 1 undercuts by only one tick, he or she obtains an expected profit *at least* equal to $0.25(101 - 99.5) = 0.375$. If dealer 1 posts the competitive price, he or she obtains a profit equal to $\frac{g(1)}{2} = 1/6 < 0.375$. Dealer 1 is better off posting a price equal to 101 rather than the competitive price. Consequently, p^* is lower than or equal to 101. This is in contrast with the case in which $\Delta = 0.3$ *and* the tick size is $g(0) = 1$, for which $p^* = 102$. For this reason, the hitting time is larger when the tick size is $g(1) = 1/3$ than when it is $g(0) = 1$. Computations yield $p^* = 100 + \frac{1}{3}$ and $\tau^* = 8$, which gives $T^*(0.3, 1/3) = 9 * 0.3 = 2.7$.

Note that a decrease in the tick size enlarges the number of possible offers on the grid.[8] Accordingly, for a given p^*, such a decrease results in a larger number of offers posted by the two dealers before the "jump" to the competitive price. This effect also contributes to the increase in the hitting time when the grid becomes

[8] For instance, there are 4 feasible prices in the interval (99.5, 103] when the grid size is $g(0) = 1$, versus 10 feasible prices when the grid size is $g(1) = 1/3$.

finer. When the tick size goes to zero, the profit obtained by posting the competitive price becomes smaller and smaller. It follows that the price at which the jump to the competitive price takes place becomes closer or even equal to the competitive price as the tick size goes to zero. At the same time, the number of feasible offers for the dealers becomes larger and larger. The combination of these two effects implies that the time it takes for the best price to reach the competitive price becomes infinite when the tick size goes to zero.

3.3. Trading Cost

Let $\tilde{TC}(\Delta, g(n))$ be the trading cost paid by the liquidity buyer. The trading cost is equal to the difference between the price paid by the liquidity buyer and the expected value of the asset. From Proposition 2, we obtain

$$\tilde{TC}(\Delta, g(n)) = \begin{cases} R_B - \tau g(n) - \mu, & \text{if } \tilde{T} \in [\tau, \tau+1) \text{ and } \tilde{T} < \tau^* + 1; \\ p(1) - \mu, & \text{if } \tilde{T} \geq \tau^* + 1. \end{cases} \tag{11}$$

Since the trading game is a zero-sum game, the expected trading cost is equal to the sum of the expected profit for the dealers; that is,

$$E(\tilde{TC}(\Delta, g(n))) = 2V(\Delta, g(n)). \tag{12}$$

COROLLARY 2. *The expected trading cost increases with dealers' reaction time.*

As explained in the previous subsection, the time it takes for the best ask price to reach the competitive level increases with dealers' reaction time. For this reason, the probability of execution, for the buy market order, at a price greater than the competitive price increases when dealers' reaction time becomes larger. It follows that the expected trading cost for the liquidity buyer increases with dealers' reaction time. This suggests that the timing of offers is a potential determinant of the trading costs. Corollary 2 predicts that, in cross-sectional analysis, controlling for the tick size and the rate of market orders arrivals, the average trading cost should be positively related to the average reaction time. This prediction could be tested using the average number of quote changes as a proxy for the average reaction time.

COROLLARY 3. *For all strictly positive values of the reaction time:*

- *The relationship between the expected trading cost and the tick size is non-monotonic.*
- *The tick size which minimizes the expected trading cost is always strictly greater than zero.*
- $E(\tilde{TC}(\Delta, 0)) \equiv \lim_{g(n)\to 0} E(\tilde{TC}(\Delta, g(n))) = R_B - \mu.$

Figure 3 illustrates this result, using the parameterization of the numerical example, for three different values of the reaction time. As can be seen in this figure,

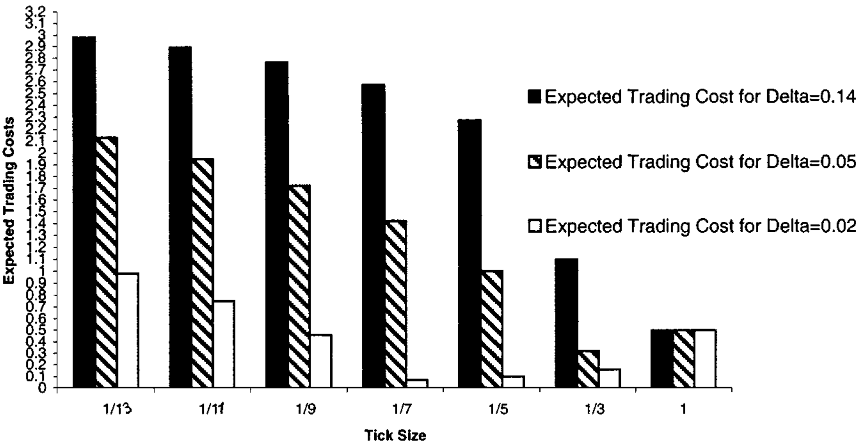

FIG. 3. Expected trading costs and tick size. This figure presents the evolution of the expected trading cost when the tick size varies from $g(0) = 1$ to $g(6) = 1/13$, for different values of the reaction time (shown on the figure). The values of the other parameters are $\mu = 99.5$, $\lambda = 1$ and $R_B = 103$.

the tick size which minimizes the expected trading cost is $g(0) = 1$ when $\Delta = 0.14$, $g(1) = 1/3$ when $\Delta = 0.05$, and $g(3) = 1/7$ when $\Delta = 0.02$.

The intuition is as follows. Consider a decrease in the tick size from $g(n)$ to $g(n+1)$. On the one hand, it reduces the wedge between the asset expected value and the competitive price. This diminishes the expected trading cost. But, on the other hand, the decrease in the tick size reduces dealers' incentive to quickly post the competitive price. As a consequence, the time during which the best ask price is greater than the competitive price lengthens (see previous subsection). It follows that the market order has a larger probability of being executed at a price above the competitive price. This effect enlarges the expected trading cost. The net impact of these two effects depends on the initial level of the tick size. For a sufficiently large initial tick size, the first effect dominates and a decrease in the tick size reduces the expected trading cost. When the tick size becomes small, the second effect dominates. In all the cases, the tick size which minimizes the expected trading cost is strictly positive. Actually, when the tick size goes to zero, the time that it takes for the best price to reach the competitive price becomes infinite (see Corollary 1).

Note that $(R_B - \mu)$ is the maximum possible value for the expected trading cost, since R_B is the liquidity buyer's reservation price. The last part of Corollary 3 states that the expected trading cost is maximum when the tick size is zero, *if* the reaction time is *strictly* positive. This result is used in the following section.

To sum up, the results of this section show that *noncompetitive spreads can be posted even though the dealers do not cooperate to set prices*. This occurs because (i) it takes some time for a dealer to react to a new offer posted by his or her competitor, and (ii) the arrival date of the market order is uncertain. The tick size and the frequency with which dealers revise their offers determine the speed with which dealers' quotes reach the competitive price and the trading cost.

4. ENDOGENOUS REACTION TIME

In this section, we endogenize dealers' reaction time in the bidding game considered in this article.

4.1. Endogenous Reaction Time with Monitoring Costs

In practice, checking the position of his or her offer relative to competitors' offers is costly for a dealer because it takes time. This monitoring cost increases with the frequency with which the dealer checks his or her quote. In our framework, it means that dealers incur a monitoring cost which decreases with the reaction time. Intuitively, this cost induces the dealers to choose a strictly positive reaction time. Although this idea is simple, it is a difficult task to endogenize the reaction time. We propose a simple extension of the model, which aims at showing that monitoring costs can explain positive reaction times, particularly when the tick size is small.

We suppose that a dealer who chooses a reaction time Δ bears a monitoring cost $C(\Delta) = cF(\Delta)$. We make the following assumptions on the monitoring cost function: (i) $\frac{\partial C}{\partial \Delta} < 0$, (ii) $C(0) > 0$, and (iii) $\lim_{\Delta \to +\infty} C(\Delta) = 0$. Note that the larger is c, the larger is the monitoring cost, other things being equal. This enables us to analyze the effect of a change in the level of the monitoring cost. We refer to c as the *monitoring cost level*.

$C(.)$ is a reduced form, which formalizes in a simple way the idea that the more frequent are the quote revisions, the higher must be the monitoring cost borne by a dealer. As a justification for our assumptions regarding $C(.)$, consider the following example. The dealers incur a cost $c(\tilde{N}) = c\tilde{N}$ if they revise $\tilde{N}$ times their offers during the trading process. In this case, it is possible to show that they expect to bear a monitoring cost $C_{ex}(\Delta)$, given by

$$C_{ex}(\Delta) = cE(\tilde{N}) = \frac{c}{e^{\lambda\Delta} - 1}. \tag{13}$$

Note that $C_{ex}(., .)$ satisfies all our assumptions. For the numerical example, we use $C_{ex}(.)$ as the monitoring cost function.[9]

We suppose that the dealers must choose their reaction time before the beginning of the trading process and that it cannot be changed thereafter. When both dealers choose a reaction time Δ, their *net* expected trading profit, in equilibrium, is

$$\Sigma(\Delta, g(n)) \equiv V(\Delta, g(n)) - C(\Delta). \tag{14}$$

Recall that $V(\Delta, g(n))$ increases and that $C(\Delta)$ decreases with dealers' reaction time. Consequently, dealers' net expected trading profit increases with dealers' reaction time.

[9] Note that if the dealers must pay a fixed cost c each time they submit a new offer, their expected cost at the beginning of the trading game is also given by $C_{ex}(.)$.

We focus on the reaction time that minimizes the expected trading cost (i.e., $V(\Delta, g(n))$), making sure that both dealers have an incentive to take part in the trading process (i.e., $\Sigma(\Delta, g(n)) \geq 0$). Let $\Delta^*(g(n), c)$ be this reaction time. It is the Nash equilibrium of the following game. The market designer proposes to the dealers a reaction time with a view to minimizing the expected trading cost. The dealers simultaneously decide to accept (enter the market) or reject (do not enter the market) this proposal. If they are indifferent, they accept. It is obvious that the market designer must propose $\Delta^*(g(n), c)$. Another approach would be to assume that the dealers simultaneously choose their reaction time. Unfortunately, we have not been able to characterize, in a tractable way, the Nash equilibrium of this game. It is very unlikely, however, that, qualitatively, the results would differ from those we obtain here.

We now study the properties of Δ^*. Recall that if $\Delta \leq \bar{\Delta}(g(n))$, convergence to the competitive price is immediate. Since $\underline{\Delta}(g(n)) > 0, \forall g(n) > 0$, it follows that $V(0, g(n)) \equiv \lim_{\Delta \to 0} V(\Delta, g(n)) = g(n)/4$. Consequently, if $C(0) \leq \frac{g(n)}{4}$, then $\Delta^*(g(n), c) = 0$. If this is not the case, the reaction time must be chosen such that the dealers just break even. This means that Δ^* is the solution of

$$\Sigma(\Delta^*, g(n)) = 0. \tag{15}$$

If this equation does not admit a solution (which can occur because $V(., g(n))$ is discontinuous), then Δ^* satisfies

$$\Sigma(\Delta^*, g(n)) < 0, \quad \text{and} \quad \Sigma(\Delta^* + \epsilon, g(n)) > 0, \quad \forall \epsilon > 0. \tag{16}$$

In this case, the market designer must propose $\Delta^* + \epsilon$ with ϵ strictly positive but very small. From now on we refer to Δ^* as the *zero expected profit reaction time*. We obtain the following result.

PROPOSITION 3. *If the monitoring cost $C(0) > \frac{g(n)}{4}$, for a given tick size, the zero expected profit reaction time is strictly positive.*

This establishes that if the tick size is small relative to the level of the monitoring cost, then the reaction time such that the dealers break even is strictly positive. The intuition is simple. When the reaction time decreases, dealers bear a larger monitoring cost. At the same time, dealers expect lower trading profits. For all reaction times lower than $\underline{\Delta}(g(n))$, the expected trading profit is equal to $\frac{g(n)}{4}$. Thus, if $C(0) > \frac{g(n)}{4}$, the reaction time such that the dealers break even is necessarily strictly positive. Figure 4 provides an illustration of the determination of the zero expected profit reaction time, in the context of the numerical example. This reaction time is such that the curves representing the functions $V(., .)$ and $C_{ex}(.)$ intersect. In this case, it can be seen that $\Delta^* \simeq 0.18$. The following proposition describes the effect of a change in the level of monitoring cost (c) both on the zero expected profit reaction time and on the expected trading cost.

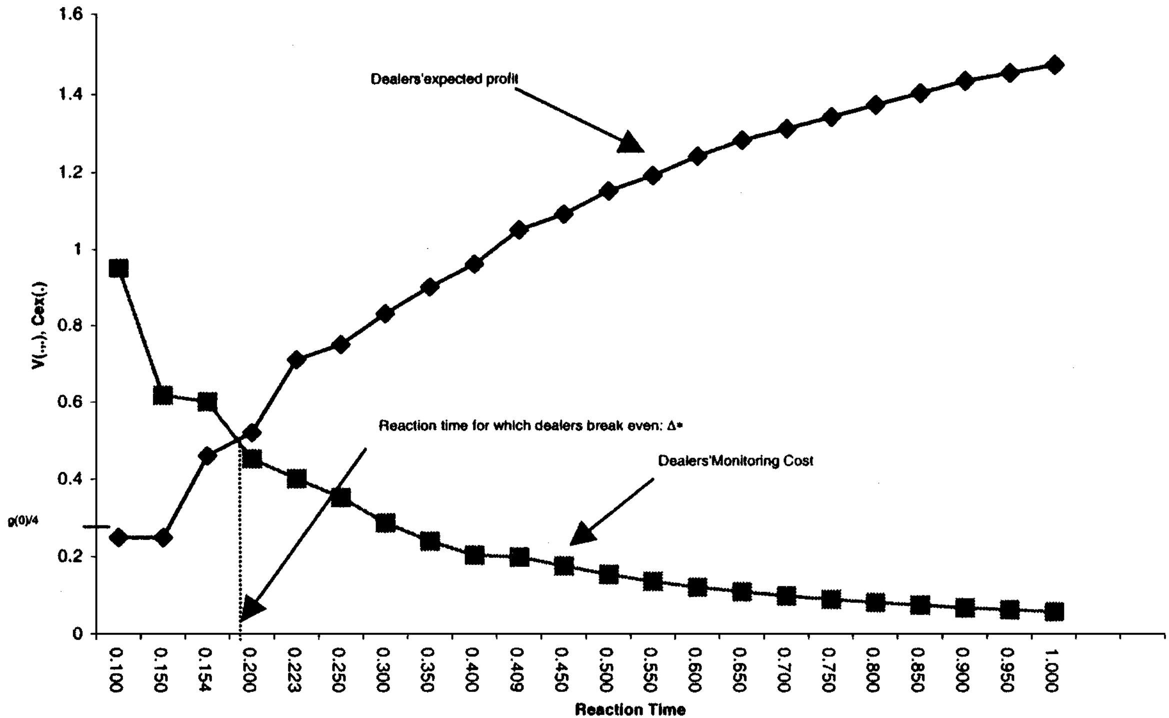

FIG. 4. Determination of the reaction time. This figure illustrates how the zero expected profit reaction time is determined. The parameter values are $\mu = 99.5$, $\lambda = 1$, $R_B = 103$, $g(0) = 1$, and $c = 0.1$. The zero expected profit reaction time is such that dealers' expected profit in the trading game is equal to dealers' monitoring cost.

PROPOSITION 4. *For a given tick size, both* (a) *the zero expected trading profit reaction time and* (b) *the expected trading cost* (*weakly*) *increase with the level of the monitoring cost.*

This result is intuitive. When c increases, the expected monitoring cost is larger, other things being equal (the curve representing $C(.)$ shifts upward in Fig. 4). This entails that dealers choose a larger reaction time (monitor the market less intensively). It follows that the hitting time is larger when the level of the monitoring cost increases. This results in a larger expected trading cost. Thus, trading costs for liquidity demanders and monitoring costs for liquidity suppliers are positively related. Note that Proposition 4 confirms the prediction associated with Corollary 2: the frequency of quote changes and the average trading cost should be negatively related.

4.2. Optimal Tick Size

We define the optimal tick size as the tick size which minimizes the expected trading cost. If the monitoring cost is strictly positive, the optimal tick size is always strictly greater than zero. Indeed, Proposition 3 shows that the reaction time for which the dealers break even is *strictly positive* when the tick size becomes sufficiently small relative to the level of the monitoring cost. It follows from the third part of Corollary 3 that $E(\tilde{TC}(\Delta^*(0), 0)) \equiv \lim_{g(n)\to 0} E(\tilde{TC}(\Delta^*(g(n)), g(n))) = R_B - \mu$. For a strictly positive tick size, convergence to the competitive level occurs in finite time. Thus, $E(\tilde{TC}(\Delta^*(g(n)), g(n))) < R_B - \mu, \forall g(n) > 0$. Consequently, the optimal tick size is bounded away from zero. In fact, the optimal tick size depends on the monitoring cost level, as established in the next proposition.

PROPOSITION 5 (Optimal Tick Size and Monitoring Cost).

- *In the presence of a monitoring cost* ($C(0) > 0$), *the tick size which minimizes the expected trading cost is always strictly greater than zero.*
- *Furthermore, the optimal tick size decreases when the monitoring cost level* (c) *decreases and converges to zero when the monitoring cost level goes to zero.*

We have already provided the explanation for the first part of this proposition. The intuition for the second part is the following. For a given tick size, when the level of the monitoring cost is large, dealers' reaction time is large as well. Now, the larger the reaction time, the larger the hitting time. Large trading costs ensue. A large tick size helps in this situation because it makes the competitive price an attractive option for the dealers. This shortens the time it takes for the best price to hit the competitive level. On the contrary, when the level of monitoring cost is small, dealers' reaction time is small. In this case, the fear of being quickly undercut induces the dealers to rapidly post the competitive price. A small tick size becomes optimal since it reduces the wedge between the competitive price and dealers' reservation price. In the limit, when the monitoring cost vanishes, the reaction time for which dealers break even goes to zero. A positive tick size is not necessary any more to speed up convergence to the competitive price. The optimal

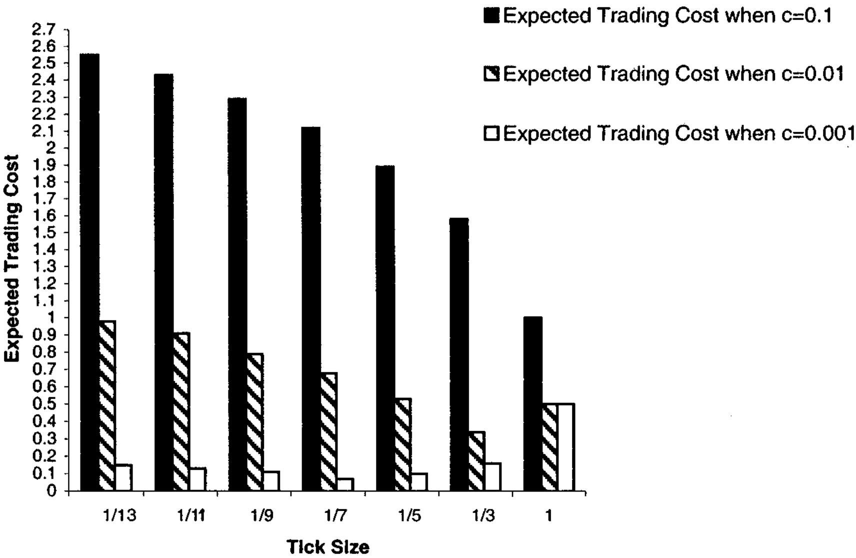

FIG. 5. Optimal tick size and monitoring cost level. This figure presents the evolution of the expected trading costs when the tick size varies from $g(0) = 1$ to $g(6) = 1/13$, for different values of the monitoring cost (shown on the figure). The values of the other parameters are $\mu = 99.5$, $\lambda = 1$, and $R_B = 103$.

tick size is zero since it eliminates the wedge between the competitive price and dealers' reservation values.

Figure 5 illustrates the result, using the parameterization of the numerical example, for three different values of the monitoring cost level. As can be seen in this figure, when $c = 0.1$, the expected trading cost is minimum when the tick size is equal to $g(0) = 1$. When $c = 0.01$, the optimal tick size is lower and equal to $g(1) = 1/3$. Finally when $c = 0.001$, a tick size equal to $g(3) = 1/7$ becomes optimal.

It is worth stressing that, for a given monitoring cost, the relationship between dealers' reaction time and the tick size is nonmonotonic. To illustrate this point, let $g^*(c)$ be the optimal tick size for a monitoring cost level c. By definition of the optimal tick size, a small increase in the tick size g, for $g < g^*(c)$, translates into a decrease in dealers' expected trading profits, for a given reaction time. Accordingly, the reaction time for which dealers break even increases with the tick size for $g < g^*(c)$. An increase in the tick size for $g > g^*(c)$ has the opposite effect on dealers' expected trading profits, for a given reaction time. Thus, dealers' reaction time must decrease with the tick size for $g > g^*(c)$. This results in an inverted U-shaped relationship between dealers' reaction time and the tick size, with the reaction time being maximum when the tick size is at its optimal level, $g^*(c)$.

Recently, the minimum price increments have decreased in various markets. For instance, Canadian exchanges implemented decimal pricing in 1996. The tick size for stocks traded under \$5 and \$10 on the AMEX has been decreased from

\$1/8 to \$1/16 (respectively in 1992 and 1995). In 1997, the NASDAQ decided to switch to quotations in 16ths. Several empirical studies[10] have shown that the decrease in the tick size in these exchanges has been followed by lower trading costs, on average. Figure 5 shows that this effect can be obtained in our model when the level of the monitoring cost decreases. For instance, if the level of the monitoring cost falls from $c = 0.1$ to $c = 0.001$, a decrease in the tick size from $g(0) = 1$ to, say, $g(2) = 1/5$ decreases the expected trading cost. Thus, our model suggests that the concomitant reduction of the tick size and the trading cost, in many exchanges, might be due to a widespread decline in the monitoring costs for liquidity providers.[11] If this interpretation is correct, the reasoning of the previous paragraph suggests that the decrease in the tick size should have triggered an increase in liquidity providers' reaction time. This prediction could be checked in future empirical work.

5. THE ROLE OF TIME PRIORITY

In this section, we consider the case in which time priority is not enforced. Instead, when there is a tie at the best price, we assume that dealers share the order equally. The evolution of the best quote in this case is qualitatively similar to the evolution obtained when time priority is enforced. Namely, the best ask price in the market converges to a focal price.[12] As a consequence, the results obtained in the previous sections concerning the roles of dealers' reaction time, the tick size, and dealers' monitoring cost still hold when time priority is not enforced. The main difference is that, when time priority is not enforced, the focal price can be larger than the competitive price. For instance, the next proposition establishes that $p(4)$ can be a focal price.

PROPOSITION 6. *If time priority is not enforced and* $R_B \geq p(4)$, *for* $\Phi \in [\frac{2}{5}, \frac{2}{3}]$, *there exists an equilibrium in which* $p(4)$ *is a focal price.*

The intuition is as follows. Consider an equilibrium in which $p(4)$ is the focal price. For $p(4)$ to be focal, it must be the case that no dealer finds it optimal to improve upon this offer. There is a strong temptation for the dealers to undercut $p(4)$, however. For instance suppose that dealer 1 is about to revise his or her offer and that $p(4)$ is the best price in the market. What is dealer 1's expected profit if he or she undercuts $p(4)$? If the liquidity trader submits his or her order before dealer 2's reaction, then dealer 1's trading profit is $p(3) - \mu = 5g(n)/2$. This is larger than dealer 1's profit if he or she does not undercut, which is $\frac{1}{2}(p(4) - \mu) = 7g(n)/4$. The order might arrive after dealer 2's reaction, however. If dealer 2 holds on his or her offer at $p(4)$ then dealer 1 is clearly better off undercutting and $p(4)$

[10] For instance, Ahn *et al.* (1996) or Bacidore (1997). See Harris (1997) for a review.

[11] This decline could be the result of the increasing automation of the trading process.

[12] The proof of this result is similar to the proof of Proposition 1. It is skipped for brevity.

cannot be focal. Thus, dealer 2 must either match or improve upon dealer 1's offer. Accordingly, suppose dealer 2 reacts by posting $p(2)$ if dealer 1 undercuts the focal price. Then, as described in the proof of Proposition 6, the best price will stay at $p(2)$ until the market order arrives. At this price, each dealer has an expected profit equal to $\frac{3g(n)}{4}$. Consequently, dealer 1's expected profit, if he or she undercuts the focal price, is $\Phi(\frac{5g(n)}{2}) + (1-\Phi)^2(\frac{3g(n)}{4})$. It follows that dealer 1 does not undercut $p(4)$ iff

$$\Phi\left(\frac{5g(n)}{2}\right) + (1-\Phi)^2\left(\frac{3g(n)}{4}\right) \leq \frac{7g(n)}{4}. \tag{17}$$

This condition is satisfied if Φ is sufficiently small (lower than $\frac{2}{3}$). Dealer 1 is deterred from undercutting $p(4)$ because he or she anticipates that such an action would trigger a price war (dealer 2 would react by undercutting) which would ultimately depress dealers' expected profits.

Other equilibria with focal prices (e.g., $p(8)$) way above the competitive price can be constructed. As these equilibria are built in the same manner as the equilibrium in which $p(4)$ is focal, we do not describe them.

Pricing strategies sustaining noncompetitive prices are often a feature of dynamic models of price competition. The interesting point here is that time priority weakens the ability of dealers to sustain noncompetitive prices. From a market design point of view, this is a nice property of time priority. To see this point, consider again the case analyzed in the previous paragraph, but assume that time priority is enforced. Dealer 1 does not hold priority at $p(4)$. In this case, dealer 1 must undercut whatever dealer 2's reaction. Actually, dealer 1 obtains no profit if he or she matches dealer 2's offer at $p(4)$. Consequently, the fear that dealer 2 might engage in a price war if he or she undercuts is not a concern for dealer 1. In this way, time priority is very conducive to undercutting.

6. A LARGE NUMBER OF DEALERS

We have assumed until now that only two dealers are posting prices. One possible concern is that the results derived in the previous sections heavily depend on this assumption. The purpose of this section is to briefly address this issue.

The driving force of our analysis is that the best price in the market does not immediately reach the competitive price if dealers' reaction time is strictly positive. We argue first that this result holds even if more than two liquidity suppliers compete for the order flow.

To make this point, let us consider a polar case in which the number of dealers is so large that each dealer has the opportunity to make only one offer. The interval of time between each dealer's offer is Δ. Notice that the possibility to make a single offer encourages dealers to post a price at which they capture time priority until the end of the bidding process. Thus, a priori, our assumption biases the model toward faster convergence than in the baseline model. Consider a dealer who is about to revise his or her offer at date τ and who observes that the best price is $p + g(n) > p(1)$. The possibilities for this dealer are either to post $p(1)$ or to post p. This dealer undercuts the best quote by only one tick iff

$$\Phi(\Delta)(p - \mu) \geq (p(1) - \mu). \tag{18}$$

This inequality is satisfied as long as $p \geq p^*$, where p^* has been defined in Proposition 2 (Eq. (18) is identical with Eq. (6)). This implies that the best price is improved by only one tick as long as it is strictly greater than p^*. Then a jump to the competitive price is observed. Thus, for a given reaction time, the evolution of the best price is exactly the same as the one obtained when two dealers move in turn repeatedly. It follows that the results of Sections 3.2 and 3.3 still apply.

Of course the reaction time itself is likely to be related to the number of dealers. Analysis of this relationship is beyond the scope of this paper, however. Intuitively we can expect the average interval of time between quote changes to decrease when the number of dealers increases, other things being equal. In practice, the number of dealers is strongly positively correlated with firm size (see, for instance, Wahal, 1997). This suggests that small and large stocks should have different quote dynamics.[13] For instance, for a given tick size, the speed of adjustment to the competitive quote must be larger for large firms. For these firms, the logic behind Proposition 5 suggests that a large tick size, as an instrument to speed up convergence to the competitive price, should be less useful. Consequently the optimal tick size must be lower for stocks that are followed by a large number of dealers. It follows that the impact of a decrease in the tick size on the expected trading cost should be greater for large firms than for small firms.[14]

7. CONCLUSIONS

In this paper, we consider a model in which dealers revise and post their offers sequentially. In this setting, when dealers are uncertain about the arrival dates for market orders, we show that it can take time for the best quote in the market to reach the competitive level. We find that the speed of convergence of the best quote to the competitive level increases with the frequency with which dealers check their offers. It follows that trading costs are related to the timing of offers. We endogenize the frequency with which dealers check their offers, assuming that monitoring the market is costly. Even for very small levels of the monitoring cost, we find that an increase in the tick size reduces the time it takes for the best quote

[13] We thank the Editor for drawing our attention to this point.

[14] Chung *et al.* (1997) analyze the impact of the decrease in the tick size on the Toronto Stock Exchange in April 1996. Consistent with our prediction, they find that the average decrease in quoted spreads has been greater for large capitalization stocks than for small capitalization stocks.

to reach the competitive level. Accordingly, a mandatory minimum price variation is a way to minimize the trading cost. We also show that the optimal tick size depends on the level of the monitoring cost.

Our model provides a framework for interpreting quote changes in between transactions. A direct testable implication of the model is that dealers should revise their offers, in between transactions, even in the absence of inventory or informational effects. The model also shows that the speed at which dealers react to a new offer (the intensity with which they monitor the market) influences the evolution of the best offer in the market in between transactions. We find that this creates a negative correlation between the frequency of quote changes and the trading costs. All these results could be considered in future empirical work. In our framework, no public information arrives during a trading round. Thus, the dealers do not face the risk of being picked off. Further research could consider the impact of this risk on dealers' monitoring decisions.

APPENDIX

Proof of Lemma 1

Consider a dealer who is about to revise his or her quote at date τ, say dealer 1. The trading history $H_{\tau-1}$ is on the equilibrium path and dealer 1 has strict priority at the best price p; that is, the state of the market is $\{p, 1\}$. It is necessarily the case that dealer 1 has chosen price p at time $\tau - 2$. Two cases can occur.

Case 1. Dealer 1 has obtained priority of execution at time $\tau - 2$ and has not been undercut at time $\tau - 1$ by his or her competitor. In equilibrium, dealer 1's competitor has posted $R(p, 0, p)$ at time $\tau - 1$. Therefore, dealer 1's belief must be $R(p, 0, p)$.

Case 2. Dealer 1 has *not* obtained priority of execution at time $\tau - 2$. Let p' be dealer 2's price just before dealer 2 revises his or her quote at time $\tau - 1$. It is the case that $p' \leq p$. Since $H_{\tau-1}$ is on the equilibrium path, then $H_{\tau-2}$ is on the equilibrium path as well. Thus, upon observing $\{p', 1\}$ at date $\tau - 1$, dealer 2 must infer correctly that dealer 1 has posted p at time $\tau - 2$. If $p' < p$, dealer 2 must react with $R(p', 1, p) = R(p, 0, p)$. Actually, dealer 2 faces exactly the same problem in state $(p, 0, p)$ and in state $(p', 1, p)$. If $p' = p$ then dealer 2 must react with $R(p, 1, p) = p$ or $R(p', 1, p) = R(p, 0, p)$. Actually, dealer 2 in state $(p, 1, p)$ can choose to keep priority at price p but otherwise he or she faces exactly the same problem as in state $(p, 0, p)$. But if $p' = p$ and $R(p, 1, p) = p$, then dealer 1 cannot have priority of execution at price p at time τ. Thus, $R(p', 1, p) = R(p, 0, p)$ and dealer 1's belief must be $R(p, 0, p)$. Q.E.D.

Proof of Proposition 1

The following two lemmata are useful in order to establish Proposition 1.

LEMMA 2. *Consider an equilibrium reaction function* $R(.)$. *If* $R(p, 0) = p$, *and if* $\{p, 1\}$ *belongs to the equilibrium path, then* $R(p, 1) = p$.

Proof. If $\{p, 1\}$ belongs to the equilibrium path, then the dealer who observes state $\{p, 1\}$ believes that his or her competitor quotes a price equal to $R(p, 0) \geq p$ (see Lemma 1). Thus, we have either $R(p, 1) = R(R(p, 0), 0)$ or $R(p, 1) = p$. Actually, a dealer in state $(p, 1, R(p, 0))$ can choose to keep priority at price p but otherwise he or she faces exactly the same problem as in state $(R(p, 0), 0, R(p, 0))$. It follows that if $R(p, 0) = p$ then $R(p, 1) = p$. Q.E.D.

LEMMA 3. *Consider an equilibrium reaction function* $R(.)$. *The set of prices such that* $R(p, 0) > p$ *and* $R(p, 1) > p$ *with* $p \geq p(1)$ *is empty.*

Proof. Assume that this is not the case and denote by p^* the largest price such that $R(p^*, 0) > p^*$ and $R(p^*, 1) > p^*$.

Step 1. Suppose that the state of the market is $\{p^*, 0\}$ and consider a dealer, say dealer 1, who revises his or her offer in this state. In equilibrium, dealer 1 must post $R(p^*, 0)$. Assume (to be contradicted) that $R(p^*, 0) > R(p^*, 1)$. Dealer 1 will renounce price priority for the next two periods, and dealer 2 will quote $R(p^*, 1)$ at the next period. Consequently, dealer 1 obtains $(1 - \Phi)^2 V(R(p^*, 1), 0)$. Now consider the following deviation for dealer 1. Dealer 1 offers $R(p^*, 1)$ instead of $R(p^*, 0)$. With this offer, dealer 1 obtains price and time priority in the next period. Actually his or her competitor does not observe this deviation since $R(p^*, 1) > p^*$ and *only the best price in the market is displayed*. Thus, dealer 1's competitor will post $R(p^*, 1)$ in the next period. Then (in two periods) dealer 1 follows the pricing policy he or she would follow in state $\{R(p^*, 1), 0\}$. This deviation gives dealer 1 $(1 - \Phi)\Phi\Pi_1(R(p^*, 1), 1, R(p^*, 1)) + (1 - \Phi)^2 V(R(p^*, 1), 0)$, which is greater than $(1 - \Phi)^2 V(R(p^*, 1), 0)$. This implies that in equilibrium, it is necessarily the case that

$$R(p^*, 0) \leq R(p^*, 1). \tag{19}$$

Step 2. Suppose that the state of the market is $\{p^*, 1\}$ and consider a dealer, say dealer 1, who revises his or her offer in this state. According to Lemma 1, dealer 1 conjectures that dealer 2 has posted $R(p^*, 0)$. Moreover, $R(p^*, 1) = R(R(p^*, 0), 0)$. Actually, the problem faced by dealer 1 is similar to the problem he or she would face in state $\{R(p^*, 0), 0\}$ because $R(p^*, 0) > p^*$. We want to prove that $R(R(p^*, 0), 0) < R(p^*, 0)$. In order to shorten the notation, let us define $p' \equiv R(p^*, 0)$.

Since $p' > p^*$, it follows from the definition of p^* that either $R(p', 1) \leq p'$ (Condition (a)) or $R(p', 0) \leq p'$ (Condition (b)).

If Condition (a) is satisfied, then we can show that $R(p', 0) < p'$ as follows. We distinguish four cases:

Case 1. $R(p', 1) = p'$. Suppose $R(p', 0) \geq p'$ in equilibrium. Then the dealer who does not have strict priority at price p' obtains zero expected profit in equilibrium. But, by undercutting his or her competitor, the dealer can obtain a strictly positive expected profit by undercutting (because $p' > p^* \geq p(1)$). Thus, $R(p', 0) \geq p'$ is not possible in equilibrium.

Case 2. $R(p', 1) \leq p' - 2g(n)$. Let us compare first the set of possible choices for a dealer who revises his or her quote in state $\{p', 1\}$ and in state $\{p' - g(n), 0\}$. The only difference is that, in the second case, the dealer cannot gain priority of execution if he or she posts p' or $p' - g(n)$. This does not prevent the dealer in state $\{p' - g(n), 0\}$ from following the same pricing policy as in state $\{p', 1\}$ since $R(p', 1) < p' - g(n)$. Thus, $R(p' - g(n), 0) = R(p', 1)$. Now assume $R(p', 0) \geq p'$ in equilibrium. Suppose dealer 1 (say) is about to revise his or her offer and the state of the market is $\{p', 0\}$. If dealer 1 posts $R(p', 0)$, he or she loses priority for the next two periods (since $R(p', 1) < p'$). Dealer 1 obtains $(1 - \Phi)^2 V(R(p', 1), 0)$. Now consider the following deviation. Dealer 1 posts $p' - g(n)$ in state $\{p', 0\}$. Since $R(p' - g(n), 0) = R(p', 1)$, dealer 1 obtains $\Phi\Pi_1(p', 1, p' - g(n)) + (1 - \Phi)^2 V(R(p', 1), 0)$. Since $p' \geq p(2)$, $\Pi_1(p', 1, p' - g(n)) > 0$ and the deviation has been profitable. A contradiction. Consequently, $R(p', 0) < p'$ in this case, as well.

Case 3. $R(p', 1) = p' - g(n) = p^*$. Assume that in equilibrium, $R(p', 0) \geq p'$. If $p^* = p' - g(n)$, then the expected profit of the dealer (say 2) moving in state $\{p', 1\}$ is bounded by $\Pi_2(p', 1, p')$. But then, $R(p', 1) = p'$ is optimal since it gives exactly this profit if $R(p', 0) \geq p'$. A contradiction to $R(p', 1) = p' - g(n)$.

Case 4. If $R(p', 1) = p' - g(n) > p^*$. Suppose that in equilibrium $R(p', 0) \geq p'$. If $R(p' - g(n), 0) < p' - g(n)$, there is a contradiction. Actually, the dealer who revises his or her quote in state $\{p', 0\}$ is better off posting $R(p' - g(n), 0)$, rather than $R(p', 0)$. Actually with the offer $R(p', 0)$, the dealer postpones profits for two periods, which is suboptimal. Consequently, in equilibrium, $R(p' - g(n), 0) \geq p' - g(n)$. But then we have $R(p' - g(n), 1) \leq p' - g(n)$ since $p' - g(n) > p^*$. Then, since $R(p' - g(n), 0) \geq p' - g(n)$ and $R(p' - g(n), 1) \leq p' - g(n)$, we can reiterate the arguments offered for Cases 1, 2, and 3 until the contradiction is found.

If Condition (b) is satisfied as an equality, then Lemma 2 implies that $R(p', 1) = p'$. But in this case, the dealer without execution priority when the best price is p' obtains a zero profit. If the dealer deviates by undercutting of at least one tick, he or she obtains a strictly positive profit (since $p' > p^* \geq p(1)$), a contradiction. Thus, for all possible cases, we have proved that $R(R(p^*, 0), 0) < R(p^*, 0)$, which means

$$R(p^*, 1) < R(p^*, 0). \tag{20}$$

But (19) and (20) cannot be simultaneously satisfied. Therefore, p^* does not exist. Q.E.D.

Now we prove Proposition 1 in two steps.

Step 1. We know from Lemma 4 that, in equilibrium, it is not possible for the reaction function to be such that $R(p(1), 1) > p(1)$ *and* $R(p(1), 0) > p(1)$. Moreover, Lemma 2 implies that $R(p(1), 1) > p(1)$ *and* $R(p(1), 0) = p(1)$ is impossible in equilibrium. Therefore, $R(p(1), 1) = p(1)$ *and* $R(p(1), 0) = p(1)$.

This proves the second part of Proposition 1.

Step 2. Now consider any price such that $p \geq p(2)$. Lemma 3 implies that at least one of those two inequalities must be true: (i) $R(p, 0) \leq p$ or (ii) $R(p, 1) \leq p$. It is straightforward that $R(p, 0) \geq p$ cannot be optimal since with such a reaction, a dealer would lose the chance to trade during two periods and would not trigger an increase in the quote of the other dealer. This implies that necessarily $R(p, 0) < p$ for $p \geq p(2)$, which is the first part of Proposition 1. Q.E.D.

Proof of Proposition 2

Consider a grid with a tick size equal to $g(n)$.

Case 1. Consider a dealer who revises his or her offer when the state of the market is $\{p(i), 0\}$, $p(i) \in [p(2), p^*]$ (p^* is defined in the proposition). If the dealer quotes $p(1)$, he or she obtains strict priority until the end of the game and an expected gain equal to $\frac{g(n)}{2}$. If the dealer undercuts the best price by one tick and if his or her order is not executed before the arrival of his or her competitor, the dealer loses priority until the end of the game, according to the conjectured equilibrium. Consequently, this choice gives the dealer an expected profit equal to $\Phi(p(i) - g(n) - \mu)$. This is lower than $\frac{g(n)}{2}$ since $p(i) \leq p^*$. Finally, if the dealer quotes a price greater than or equal to p, his or her competitor maintains his or her quote at the same level as long as he or she is not undercut or his or her competitor quotes $p(1)$ at the next round. Consequently, the dealer must choose $p(1)$, which proves item 2 of Proposition 2.

Case 2. Now consider a dealer who revises his or her offer in state $\{p(i), 0\}$ with $p(i) \in [p^* + g(n), R_B]$. In this case, if the dealer undercuts by only one tick, he or she obtains *at least* $\Phi(p - g(n) - \mu)$, which is larger than $g(n)/2$ since $p(i) > p^*$. Moreover, the dealer's competitor will not change his or her offer as long as this offer has not been undercut since $p(i) > p^*$. Consequently, in this case $R(p(i), 0) = p(i) - g(n)$ is a best response. This proves item 1 of Proposition 2.

Case 3. Consider a dealer who revises his or her offer in state $\{p(i), 1\}$ with $p(i) \in [p^*, R_B]$. Proposition 1 implies that $\{p, 1\}$ is never on the equilibrium path if $p(1) < p < R_B$. Consequently, if $p(1) < p < R_B$, the dealer believes that his or her competitor quotes $p(i) + g(n)$. If $p = R_B$, everything is as if the dealer were in state $\{R_B + g(n), 0\}$. Since $p(i) + g(n) \geq p^* + g(n)$, using the argument developed in the previous case, it is straightforward that $R(p(i), 1) = p(i)$ is a best response for the dealer. This proves item 3 of Proposition 2.

Case 4. Finally, consider a dealer in state $\{p(i), 1\}$, with $p(i) \in [p(1), p^* - g(n)]$. For the same reason as in Case 3, the dealer believes that his or her competitor has offered $p(i)+g(n)$. Since $p(i)+g(n)\le p^*$, using the argument developed in Case 1, it is straightforward that $R(p(i),1)=p(1)$ is a best response for the dealer. This proves item 4 of Proposition 2. Q.E.D.

Proof of Corollary 2

When time priority is enforced, if $\Delta \le \overline{\Delta}(g(n))$, the best price converges immediately to the competitive price. This implies

$$E(\tilde{TC}(\Delta, g(n))) = \frac{g(n)}{2} \quad \forall \Delta < \overline{\Delta}(g(n)). \tag{21}$$

Now if $\Delta > \overline{\Delta}(g(n))$, it takes time to converge, and the expected trading profit for a dealer is given by

$$V(\Delta, g(n)) = \frac{1}{2}\left(\Phi\left[\sum_{\tau=0}^{\tau=\tau^*(\Phi)} (1-\Phi)^{\tau}\left(a_{\tau}^{m}-\mu\right)\right] + (1-\Phi)^{\tau^*(\Phi)+1}(p(1)-\mu)\right), \tag{22}$$

where we emphasize the dependence between τ^* and Φ for the sake of the proof. Using the definition of τ^*, it is direct that τ^* is (weakly) increasing in Φ. It is then simple to show that the RHS of Eq. (22) is increasing with Φ. Since Φ increases with Δ, it follows that $V(\Delta, g(n))$ increases with Δ when $\Delta > \overline{\Delta}(g(n))$. Since $E(\tilde{TC}(\Delta, g(n))) = 2V(\Delta, g(n))$, the corollary is proved. Q.E.D.

Proof of Corollary 3

Take Δ as given. Consider a tick size $g(n')>0$, small enough so that $\Delta > \overline{\Delta}(g(n'))$. If $\Delta > 0$, such a tick size exists since $\overline{\Delta}(g(n))$ increases with $g(n)$ and $\overline{\Delta}(0)=0$. From the definition of $\overline{\Delta}(g(n))$, it follows that for all tick sizes $g(n)<g(n')$, $\tau^*(g(n))>0$. Consequently, for all tick sizes $g(n)<g(n')$, we get

$$V(\Delta, g(n)) = \frac{1}{2}\left(\Phi\left[\sum_{\tau=0}^{\tau=\tau^*(g(n))} (1-\Phi)^{\tau}\left(a_{\tau}^{m}-\mu\right)\right] \right.$$
$$\left. + (1-\Phi)^{\tau^*(g(n))+1}(p(1)-\mu)\right). \tag{23}$$

Since $a_{\tau}^{m} = R_B - \tau g(n)$ for $\tau \le \tau^*$, this can be rewritten as

$$V(\Delta, g(n)) = \frac{1}{2}\left(\Phi\left[\sum_{\tau=0}^{\tau=\tau^*(g(n))} (1-\Phi)^{\tau}(R_B - \tau g(n) - \mu)\right] \right.$$
$$\left. + (1-\Phi)^{\tau^*(g(n))+1}(g(n)/2)\right). \tag{24}$$

Using the fact that $\lim_{g(n)\to 0} \tau^*(g(n)) = \infty$, we get

$$V(\Delta, 0) \equiv \lim_{g(n)\to 0} V(\Delta, g(n)) = \frac{R_B - \mu}{2} \quad \forall \Delta > 0. \tag{25}$$

Recall that $E(\tilde{TC}(\Delta, g(n))) = 2V(\Delta, g(n))$. Consequently,

$$E(\tilde{TC}(\Delta, 0)) = 2V(\Delta, 0) = R_B - \mu \quad \forall \Delta > 0. \tag{26}$$

Note that the trading cost cannot be larger than $R_B - \mu$ since R_B is the liquidity buyer's reservation price. Convergence to the competitive price takes place in finite time if $g(n) > 0$. Thus it is immediate that $E(\tilde{TC}(\Delta, g(n))) < E(\tilde{TC}(\Delta, 0))$ $\forall g(n) > 0$ if $\frac{g(n)}{2} < R_B - \mu$, that is, if $R_B > p(1)$, which is a necessary condition for trading to take place. This proves that the tick size which minimizes the expected trading cost must be strictly positive. Q.E.D.

Proof of Proposition 3

If $C(0) > \frac{g(n)}{4}$, then $\Sigma(0, g(n)) < 0$. On the other hand, $\lim_{\Delta\to+\infty} C(\Delta) = 0$, which implies that $\lim_{\Delta\to+\infty} \Sigma(\Delta, g(n)) > 0$ (because $V(\Delta, g(n)) > 0 \, \forall \Delta \geq 0$). This implies that $\Delta^*(g(n), c)$ exists and is strictly positive if $C(0) > \frac{g(n)}{4}$. Q.E.D.

Proof of Proposition 4

The first part of the proposition is straightforward. For the second part, consider two levels of monitoring costs, c_h and c_l, with $c_h > c_l$. From the first part of Proposition 4, we know that $\Delta^*(g(n), c_h) \geq \Delta^*(g(n), c_l)$. Since $E(\tilde{TC}(\Delta, g(n))$ increases with Δ, we have $E(\tilde{TC}(\Delta^*(g(n), c_h), g(n))) \geq E(\tilde{TC}(\Delta^*(g(n), c_l), g(n)))$. Q.E.D.

Proof of Proposition 5

We recall that $\bar{\Delta}(g(n)) = \frac{-1}{\lambda} \ln(1 - \frac{g(n)}{2}(R_B - \mu)^{-1})$. Let $\mathcal{S}(c)$ be the set of tick sizes such that $\Delta^*(g(n), c) \leq \bar{\Delta}(g(n))$. Suppose that c is sufficiently small for $g(0) \in \mathcal{S}(c)$. We know from Corollary 1 that $\forall g(n) \in \mathcal{S}(c)$, the adjustment of the best price to the competitive level is immediate, and dealers' expected profit is equal to $\frac{g(n)}{4}$. Thus, $\forall g(n) \in \mathcal{S}(c)$, Δ^* solves

$$\frac{g(n)}{4} = C(\Delta^*). \tag{27}$$

Consequently, $\forall g(n) \in \mathcal{S}(c)$, $\Delta^*(g(n), c)$ decreases with $g(n)$. Furthermore, $\bar{\Delta}$ increases with $g(n)$. This implies that if $g(n') \in \mathcal{S}(c)$ and $g(n) > g(n')$, then $g(n) \in \mathcal{S}(c)$ as well. Thus, there exists $n^*(c)$ such that $\mathcal{S}(c) = [g(n^*(c)), g(0)]$. It is clear that $g(n^*(c))$ is an upper bound for the optimal tick size. We now prove that $g(n^*(c))$ decreases with c. Consider two levels of monitoring costs, c_h and c_l, with $c_h > c_l$.

By definition,

$$\bar{\Delta}(g(n^*(c_h)) > \Delta^*(g(n^*(c_h)), c_h). \tag{28}$$

As Δ^* increases with c, the previous inequality entails

$$\bar{\Delta}(g(n^*(c_h)) > \Delta^*(g(n^*(c_h)), c_l). \tag{29}$$

This implies that $\mathcal{S}(c_h) \subset \mathcal{S}(c_l)$; i.e., $g(n^*(c_h)) \geq g(n^*(c_l))$. Thus the optimal tick size decreases when c decreases since $g(n^*(c))$ is an upper bound on the optimal tick size.

Finally, suppose (to be contradicted) that $\lim_{c \to 0} g(n^*(c)) = g' > 0$. Thus g' is such that $\forall g(n) < g'$ and for c sufficiently small, $\Delta^*(g(n), c) > \bar{\Delta}(g(n))$. But consider a tick size $g(n) < g'$. We can always choose c sufficiently small so that $V(\bar{\Delta}(g(n)), g(n)) - C(\bar{\Delta}(g(n))) > 0$. But this implies that $\Delta^*(g(n), c) \leq \bar{\Delta}(g(n))$ since $\Sigma(\Delta, g(n))$ increases with Δ. A contradiction. Thus, $\lim_{c \to 0} g(n^*(c)) = 0$. Q.E.D.

Proof of Proposition 6

Assume $R_B \geq p(4)$ and $\Phi \in [\frac{2}{5}, \frac{2}{3}]$. Consider the following reaction function:

1. $R(p, 0) = p - g$ if $p^{**} < p \leq R_B$.
2. $R(p, 1) = p$ if $p^{**} \leq p \leq R_B$.
3. $R(p, 0) = p(4)$ if $p(4) \leq p \leq p^{**}$.
4. $R(p, 1) = p(4)$ if $p(4) \leq p < p^{**}$.
5. $R(p(3), 0) = p(2)$ and $R(p(3), 1) = p(4)$.
6. $R(p(2), 0) = R(p(2), 1) = p(2)$.

Here $p^{**} = p(5)$ if $\Phi \in [\frac{3}{7}, \frac{2}{3}]$ and $p^{**} = p(6)$ if $\Phi \in [\frac{2}{5}, \frac{3}{7}]$.

Computations similar to the computations used in the proof of Proposition 2 show that this reaction function is an equilibrium when time priority is not enforced. We skip the details for brevity. It is direct that $p(4)$ is the focal price associated with this reaction function. We now explain why we need conditions on Φ in this case. First, as explained in the first paragraph following Proposition 6, we need $\Phi \leq \frac{2}{3}$ in order to prevent a dealer from undercutting the focal price. Second, if a dealer, say dealer 1, were to undercut the focal price, we would need dealer 2 to retaliate. This means $R(p(3), 0) = p(2)$. However, dealer 2 could decide not to react and choose $R(p(3), 0) = p(4)$. Since $R(p(3), 1) = p(4)$, this would bring back the best price at $p(4)$ when dealer 1 revises his or her offer. Thus, for $R(p(3), 0) = p(2)$ to be optimal, we need

$$\Phi(p(2) - \mu) + (1 - \Phi)\frac{(p(2) - \mu)}{2} \geq (1 - \Phi)\left(\frac{p(4) - \mu}{2}\right). \tag{30}$$

This is the case if $\Phi \geq \frac{2}{5}$. Q.E.D.

REFERENCES

Ahn, H. J., Cao, C., and Choe, H. (1996). Tick size, spread and volume, *J. Finan. Intermed*. **5**, 2–22.

Angel, J. (1997). Tick size, share prices and stock splits, *J. Finance* **52**, 655–681.

Anshuman, V. R., and Kalay, A. (1998). Market-making with discrete prices, *Rev. Finan. Stud.* **11**, 81–109.

Bacidore, J. (1997). The impact of decimalization on market quality: An empirical investigation of the Toronto Stock Exchange, *J. Finan. Intermed*. **6**, 92–120.

Bernhardt, D., and Hughson, E. (1996). Discrete pricing and the design of dealerships markets, *J. Econ. Theory* **71**, 148–182.

Biais, B., Hillion, P., and Spatt, C. (1995). An empirical analysis of the limit order book and the order flow in the Paris Bourse, *J. Finance* **50**, 1655–1689.

Chordia, T., and Subrahmanyam, A. (1995). Market-making, the tick size and payment for the order flow, *J. Bus.* **68**, 543–575.

Chung, R., Kryzanowski, L., and Zhang, H. (1997). Decimalization's winners and losers, *Can. Investment Rev.* **Winter**, 35–39.

Domowitz, I. (1993). A taxonomy of automated trade execution systems, *J. Int. Money Finance* **12**, 607–631.

Dutta, P., and Madhavan, A. (1997). Competition and collusion in dealer markets, *J. Finance* **52**, 245–276.

Easley, D., Kiefer, N., O'Hara, M., and Paperman, J. (1996). Liquidity, information and infrequently traded stocks, *J. Finance* **51**, 1405–1436.

Garman, M. (1976). Market microstructure, *J. Finan. Econ.* **3**, 257–275.

Gouriéroux, C., Jasiak, J., and LeFol, G. (1997). "Intra-day Market Activity," working paper, CREST.

Harris, L. (1997). "Decimalization: A Review of the Arguments and Evidence," mimeo, University of Southern California.

Kandel, E., and Marx, L. (1997). NASDAQ market structure and spread patterns, *J. Finan. Econ.* **45**, 61–89.

Maskin, E., and Tirole, J. (1988). A theory of dynamic oligopoly. II. Price competition, kinked demand curves and Edgeworth cycles, *Econometrica* **56**, 571–599.

Wahal, S. (1997). Entry, exit, market makers and the bid–ask spread, *Rev. Finan. Stud.* **10**, 871–901.