Limit Order Book as a Market for Liquidity

Thierry Foucault
HEC School of Management, Paris and CEPR

Ohad Kadan
John M. Olin School of Business, Washington University in St. Louis

Eugene Kandel
School of Business Administration, and Department of Economics, Hebrew University

We develop a dynamic model of a limit order market populated by strategic liquidity traders of varying impatience. In equilibrium, patient traders tend to submit limit orders, whereas impatient traders submit market orders. Two variables are the key determinants of the limit order book dynamics in equilibrium: the proportion of patient traders and the order arrival rate. We offer several testable implications for various market quality measures such as spread, trading frequency, market resiliency, and time to execution for limit orders. Finally, we show the effect of imposing a minimal price variation on these measures.

The timing of trading needs is not synchronized across investors, yet trade execution requires that counterparties trade simultaneously. Markets address this problem in one of the three ways: call auctions, dealer markets, and limit order markets. Call auctions require participants to either wait or trade ahead of their desired time; no one gets immediacy, unless by chance. Dealer markets, on the contrary, provide immediacy to all at the same price, whether it is desired or not. Finally, a limit order market allows investors to demand immediacy, or supply it, according to their choice. The growing importance of limit order markets suggests that this feature is valuable, which in turn implies that traders value order

We thank Bruno Biais, David Easley, Larry Glosten, Larry Harris, Joel Hasbrouck, Frank de Jong, Pete Kyle, Leslie Marx, Narayan Naik, Maureen O'Hara (the editor), Christine Parlour, Patrik Sandas, Duane Seppi, Ilya Strebulaev, Isabel Tkach, Avi Wohl, and two anonymous referees for very helpful comments and suggestions. Comments by seminar participants at Amsterdam, BGU, Bar Ilan, CREST, Emory, Hebrew, Illinois, Insead, LBS, Stockholm, Tel Aviv, Thema, Toulouse, Wharton, and by participants at the Western Finance Association 2001 meeting, the CEPR 2001 Symposium at Gerzensee, and the RFS 2002 Imperfect Markets Conference have been very helpful as well. The authors thank the J. Nachmias Fund, and The Kruger Center at Hebrew University for financial support. Thierry Foucault is also affiliated with GREGHEC. He thanks la Fondation HEC for additional financial support. Finally, we are grateful to Michael Borns for expert editorial assistance. Address correspondence to Eugene Kandel, Department of Economics, Hebrew University, Mt. Scopus, Jerusalem, Israel, 91905", and CEPR, or e-mail:mskandel@mscc.huji.ac.il.

doi:10.1093/rfs/hhi029 Advance Access publication August 31, 2005

execution speed differently.[1] In this article, we explore a dynamic model of limit order trading in which traders differ in their level of impatience.

Limit and market orders constitute the core of any continuous limit order trading system. A market order guarantees immediate execution at the best price available upon the order arrival. It represents demand for the immediacy of execution. With a limit order, a trader can improve the execution price relative to the market order price, but the execution is neither immediate nor certain. A limit order represents supply of immediacy to other traders. The optimal order choice ultimately involves a trade-off between the cost of delayed execution and the cost of immediacy. This trade-off was first suggested by Demsetz (1968, p. 41), who states: "*Waiting costs are relatively important for trading in organized markets, and would seem to dominate the determination of spreads.*" He argued that more aggressive limit orders would be submitted to shorten the expected time to execution, driving the book dynamics.[2]

Building on this idea, we study how traders' impatience affects order placement strategies, bid-ask spread dynamics, and market *resiliency*. Black (1971) and Kyle (1985) define a liquid market as being (a) tight—small spreads, (b) deep—small price impact, and (c) resilient—prices recover quickly after liquidity demand shocks [see also Harris (2003), p. 400]. The determinants of spreads and market depth have been extensively analyzed. In contrast, market resiliency, an inherently dynamic phenomenon, has received little attention in theoretical research. Yet, mean reversion in the spread following liquidity shocks is an important feature of data on order driven markets.[3] Our dynamic equilibrium framework allows us to fill this gap.

The model features buyers and sellers arriving sequentially. All are liquidity traders who would like to buy/sell one unit regardless of the prevailing price. Traders differ in their costs of delaying execution: they arrive randomly as either patient (low waiting cost) or impatient (high waiting cost). Each trader chooses to place a market or a limit order, conditional on the state of the book, so as to minimize his total execution cost which includes the cost of waiting. Under several simplifying assumptions, we derive (i) the equilibrium order placement strategies,

[1] Jain (2002) shows that in the late 1990s, 48% of the 139 stock markets throughout the world are organized as a pure limit order book, while another 14% are hybrid with the limit order book as the core engine. Examples of limit order markets include Island and Euronext in equity markets, Reuters D-2002 in the FX market or MTS in the bond market.

[2] Demsetz (1968) focuses on the NYSE. This market is not a pure limit order market since liquidity is supplied both by limit order traders and dealers (the specialists). Demsetz sees waiting costs as particularly important for traders who choose between limit and market orders.

[3] For instance, Biais, Hillion and Spatt (1995) find that liquidity demand shocks, manifested by a sequence of market orders, raise the spread. Then the spread reverts to the competitive level as liquidity suppliers place new orders within the prevailing quotes. DeGryse *et al.* (2003) and Coppejans, Domowitz, and Madhavan (2003) have also studied this phenomenon.

(ii) the expected time to execution for limit orders, (iii) the stationary probability distribution of the spread, and (iv) the expected duration between trades (conditional on the size of the inside spread). In equilibrium, patient traders tend to provide liquidity to less patient traders.

A string of market orders, that is, a liquidity shock, increases the spread. We *measure* market resiliency by the probability that, after a liquidity shock, the spread reverts to its former level before the next transaction. Factors that induce traders to post-aggressive limit orders make the market resilient. For instance, other things being equal, an increase in the proportion of patient traders reduces liquidity demand and lengthens the expected time to execution of limit orders. As a result, liquidity suppliers submit more aggressive limit orders to reduce their waiting times, in line with Demsetz's (1968) intuition. Consequently, when the proportion of patient traders increases, the spread narrows more rapidly, making the market more resilient. Higher arrival rates translate into shorter waiting times for limit order traders. The latter respond with less aggressive limit orders, thus more orders are required before the spread reverts to its competitive level; consequently, market resiliency *decreases* in the order arrival rate.

Interestingly, the distribution of spreads depends on the composition of the trading population. Large spreads are more frequent in markets dominated by impatient traders, *because* these markets are less resilient. A reduction of the tick size in these markets can result in higher spreads. Actually, this reduction impairs market resiliency by enabling traders to bid less aggressively.

We derive several empirical predictions.[4] The advent of high frequency databases has spurred an interest in the role of time in the trading process [e.g., Easley and O'Hara (1992), Engle and Russel (1998), Hasbrouck (1999), and Lo, McKinlay, and Zhang (2001)]. The time between trades in our model is endogenous since a transaction occurs when a trader opts for a market order. We show that the average time until a transaction, conditional on the quoted spread for the prior transaction, increases with the size of the spread. Furthermore, there is a positive relation between this conditional duration and market resiliency. This result stems from the fact that, other things being equal, both market resiliency and the expected duration between trades decrease with the proportion of impatient traders. We also propose to explain intraday liquidity patterns by time-series variations in the proportion of patient traders. Assuming traders become more impatient over the course of the trading day, our model predicts an increase in spreads and trading frequency toward the

[4] The number of empirical papers on limit order markets is growing fast. These include Handa and Schwartz (1996), Harris and Hasbrouck (1996), Kavajecz (1999), Sandås (2001), Hollifield, Miller, and Sandås (2004), Hollifield *et al*. (2003), Kavajecz and Odders-White (2003) and other references which are mentioned later in the article.

end of the day. This should be concomitant with a decline in limit order aggressiveness and market resiliency. Whereas the first two predictions are consistent with the empirical findings, as far as we know the latter has not yet been tested.

Most of the models in the theoretical literature such as Glosten (1994), Chakravarty and Holden (1995), Rock (1996), Seppi (1997), Biais, Martimort, and Rochet (2000), or Parlour and Seppi (2003) focus on the optimal bidding strategies for limit order traders. These models are static, which precludes the analysis of the determinants of market resiliency. Furthermore, the choice between market and limit orders is exogenous in these models. In particular, this choice is not explicitly related to the level of waiting costs, as it is in our model.[5]

Parlour (1998), Foucault (1999), and Goettler, Parlour, and Rajan (2003) study dynamic models.[6] Parlour (1998) shows how the order placement decision is influenced by the depth available at the inside quotes. Foucault (1999) analyzes the impact of the risk of being picked off on traders' order placement strategies. Goettler, Parlour, and Rajan (2003) model limit order trading as a stochastic sequential game and develop a technique to solve for the equilibrium numerically. In these models, limit order traders do not bear waiting costs, and time to execution does not influence traders' bidding strategies.[7] In contrast, time to execution plays a central role here.

There is no asymmetric information among traders in our model. This approach seems reasonable, as a first cut, since it is very difficult to solve dynamic models with traders who can strategically choose between market and limit orders. Frictions in our model (the bid-ask spread and the waiting time) are due to (i) the waiting costs and (ii) strategic rent-seeking by patient traders. Frictions that are not caused by informational asymmetries appear to be large in practice [see Huang and Stoll (1997) or Madhavan, Richardson, and Roomans (1997)]. For instance, Huang and Stoll (1997) estimate that 88.8% of the bid-ask spread on average is due to non-informational frictions (so-called order processing costs). Given this evidence, it is important to understand the theory of price formation when frictions are not due to informational asymmetries.

The article is organized as follows. Section 1 describes the model. Section 2 derives the equilibrium of the limit order market and analyzes the determinants of market resiliency. Section 3 discusses in detail the

[5] In extant models, traders who submit limit orders may be seen as very patient, while those who submit market orders may be seen as extremely impatient. We consider a less polar case.

[6] Angel (1994), Domowitz and Wang (1994), and Harris (1998) study models with an exogenous order flow. We use more restrictive assumptions on the primitives of the model that anable us to endogenize the order flow. Rosu (2004) uses a similar approach."

[7] In Parlour (1998) traders' utility does not depend on their execution timing *during* the market day, i.e., there is no cost of waiting.

empirical implications of the model. In Section 4, we explore the effect of a change in tick size on measures of market quality, and Section 5 addresses robustness issues. Section 6 concludes and discusses the limitations of our approach. All proofs are in the Appendix.

1. Model

Consider a continuous market for a single security, organized as a limit order book without intermediaries. We assume that latent information about the security value determines the range, $[B, A]$, of admissible prices. Specifically, a competitive fringe of traders stands ready to sell and buy an unlimited number of shares at prices A and B ($A > B > 0$), respectively. We assume that A and B are constant over time, thus all the prices in the limit order book stay in the range $[B, A]$.[8] The goal of this model is to investigate price dynamics within this interval; these are determined by the order submission strategies followed by the traders.

All prices and spreads, but not waiting costs and traders' valuations, are placed on a discrete grid, and are expressed as a multiple of the tick size, denoted by $\Delta > 0$. The inside spread is $s \equiv a - b$, where a and b are the best ask and bid quotes in the market, expressed in number of ticks. By construction, $a \leq A$, $b \geq B$, and $s \leq K \equiv A - B$. Occasionally, we express prices and spreads in monetary terms, rather than in number of ticks, using a superscript "m", for example $s^m = s\Delta$.[9] We omit time subscripts on variables since we focus on stationary equilibria.

1.1 Timing

This is an infinite horizon model with a continuous time line. Traders arrive at the market according to a Poisson process with parameter $\lambda > 0$: the number of traders arriving during a time interval of length τ is distributed according to a Poisson distribution with parameter $\lambda\tau$. As a result, the inter-arrival times are distributed exponentially, and the expected time between arrivals is $\frac{1}{\lambda}$. We refer to the time elapsed between two consecutive trader arrivals as a *period*.

1.2 Patient and impatient traders

Traders are risk neutral. Each trader arrives as either a buyer or a seller for one share of security. Let V_{buyer} and V_{seller} be buyers' and sellers' valuations. In order to justify our classification to buyers and sellers, we assume that $V_{buyer} > A\Delta$ and $V_{seller} < B\Delta$. Upon arrival, traders observe the limit order book and must submit an order. They can submit either (i) a market

[8] Seppi (1997) and Parlour and Seppi (2003) use a similar specification of the admissible price range.

[9] For instance, $s = 4$ means that the spread is equal to four ticks. If the tick is equal to $\$\frac{1}{16}$, then the corresponding spread expressed in dollars is $s^m = \$0.25$.

order, which gets immediate execution at the best quote or (ii) a limit order, which results in a better execution price, but delays execution.

Traders bear waiting costs that are proportional to the amount of time elapsed between their arrival and the completion of their transaction. Hence, agents face a trade-off between the execution price and the time to execution. Traders are not required to execute their trade by a fixed deadline but they cannot choose not to trade [as in Admati and Pfleiderer (1988) for instance].

Both buyers and sellers can be of two types which differ by the magnitude of their waiting costs. Type *P* traders are relatively patient and incur a waiting cost of δ_P *per unit of time*. Type *I* traders are relatively impatient and incur a waiting cost of δ_I, where $\delta_I \geq \delta_P \geq 0$. The proportion of patient traders in the population is θ_P ($1>\theta_P>0$), and the proportion of impatient traders is $\theta_I = 1 - \theta_P$. These proportions remain constant over time, and the arrival process is independent of the type distribution.

A patient trader represents, for example, a portfolio manager rebalancing his portfolio due to considerations of long-term fundamental value (a "value trader"). In contrast, arbitrageurs, technical traders, or indexers, who seek to mimic the return on a specific stock or index, are impatient traders. Keim and Madhavan (1995) provide evidence supporting this interpretation. They find that indexers and technical traders are more likely to place market orders, while value traders in general place limit orders. Brokers executing agency trades are also impatient traders, since waiting may result in a worse price and therefore could lead to claims of negligence.

1.3 Trading mechanism

Limit orders are stored in the limit order book and are executed in sequence according to *price priority*. We make the following simplifying assumptions about the market structure.

Assumption A.1. Each trader arrives only once, submits a market or a limit order and exits. Submitted orders cannot be canceled or modified.

Assumption A.2. Limit orders must be price improving, that is, narrow the spread by at least one tick.

Assumption A.3. Buyers and sellers alternate with certainty, for example, first a buyer arrives, then a seller, then a buyer, and so on. The first trader is a buyer with probability 0.5.

Assumptions A.1–A.3 facilitate the analysis of the trading game, primarily for two reasons. First, they enable us to solve for the equilibria by induction (see Section 2.1 for a detailed explanation). Second, they imply that the order placement strategies depend only on the inside spread (and not on all the orders in the book). In Section 5, we demonstrate using examples that the main implications and the economic intuitions of the model persist when assumptions A.2 and A.3 are relaxed. We also explain why relaxation of these assumptions increases the complexity of the problem in a way that precludes a general analytical solution. Finally, we discuss in Section 5 the limitations imposed by Assumption A.1, namely that traders cannot cancel and resubmit their limit orders.

1.4 Order placement strategies

Let p_{buyer} and p_{seller} be the execution prices of buyers and sellers, respectively. A buyer either (i) submits a market order and pays the lowest ask a or (ii) submits a limit buy order which narrows the spread. Similarly, a seller either receives the largest bid b or submits a limit sell order. The execution prices can be expressed as

$$p_{buyer} = a - j;\ p_{seller} = b + j \text{ just } j \in \{0, \ldots, s-1\},$$

with $j = 0$ for a market order and $j > 0$ for a limit order creating a spread of size j. Recall that $s \equiv a - b$ is the inside spread *prior* to the order arrival. It is convenient to use j, rather than p_{buyer} or p_{seller}, as the trader's decision variable. We say that a trader uses a "*j-limit order*" when he posts a limit order that creates a spread of j ticks. The expected time to execution of a j-limit order is denoted by $T(j)$. Thus, the expected waiting cost of a j-limit order is $\delta_i T(j),\ i \in \{P, I\}$. As a market order entails immediate execution, we have $T(0) = 0$.

The expected profit of trader i $(i \in \{P, I\})$ who submits a j-limit order is:

$$\Pi_i(j) = \begin{cases} V_{buyer} - p_{buyer}\Delta - \delta_i T(j) = \left(V_{buyer} - a\Delta\right) + j\Delta - \delta_i T(j) & \text{for a buyer} \\ p_{seller}\Delta - V_{seller} - \delta_i T(j) = (b\Delta - V_{seller}) + j\Delta - \delta_i T(j) & \text{for a seller.} \end{cases}$$

Expressions in parenthesis represent profits associated with market order submission. These profits are determined by a trader's valuation and the best quotes in the market. It immediately follows that trader i $(i \in \{P, I\})$ observing the spread s chooses optimally the order that solves:

$$\max_{j \in \{0, \ldots s-1\}} \pi_i(j) \equiv j\Delta - \delta_i T(j), \tag{1}$$

for buyers and sellers alike. An order placement strategy for trader i is a mapping, $o_i(\cdot)$, that assigns a j-limit order, $j \in \{0, \ldots, s-1\}$, to every possible spread $s \in \{1, \ldots, K\}$. If a trader is indifferent between two limit orders with different prices, we assume that he submits the limit order creating the larger spread.

Equilibrium definition. An equilibrium of the trading game is a pair of order placement strategies, $o_P^*(\cdot)$ and $o_I^*(\cdot)$, such that the orders prescribed by the strategies solve Program (1) when the expected waiting time, $T^*(\cdot)$, is computed assuming that traders follow strategies $o_P^*(\cdot)$ and $o_I^*(\cdot)$.

Traders' optimal order placement strategies depend on the expected waiting time function. In turn, the waiting time function is endogenous and is determined by traders' order placement strategies. We will show that the equilibrium waiting time function, $T^*(j)$, is non-decreasing in *j*; thus, traders face the following trade-off: a better execution price (larger value of *j*) can only be obtained at the cost of a larger expected waiting time. Finally, notice that we restrict our attention to *stationary* order placement strategies and waiting time functions (i.e., the strategies and waiting times do not depend on the time at which the order is submitted). Hence, we only focus on the stationary equilibria of the trading game analyzed in this article. This restriction is natural because all exogenous parameters are assumed to be stationary.

2. Equilibrium Order Placement Strategies and Market Dynamics

In this section, we first characterize the equilibrium order placement strategies. Then, we study how spreads evolve in between transactions and analyze the determinants of market resiliency. We identify three different patterns for the dynamics of the limit order book: (a) *strongly resilient*, (b) *resilient*, and (c) *weakly resilient*. The pattern is determined by the characteristics of the traders' population: (i) the proportion of patient traders and (ii) the difference in waiting costs between patient and impatient traders. These parameters also determine traders' bidding aggressiveness and the resulting stationary distribution of spreads.

2.1 Expected waiting time

Consider a trader who chooses a *j*-limit order. Let $\alpha_k(j)$ be the probability that the next trader responds with a *k*-limit order, $k \in \{0, 1, \ldots, j-1\}$ ($k = 0$ stands for a market order). Lemma 1 characterizes the expected waiting time for the order placed by the first trader as a function of the next trader's order placement strategy [described by the $\alpha_k(j)$]:

Lemma 1. *The expected waiting time for the execution of a j-limit order is*:

- $T(j) = \dfrac{1}{\lambda}$ if $j = 1$,
- $T(j) = +\infty$ if $\alpha_0(j) = 0$ and $j \in \{2, \ldots, K-1\}$,
- $T(j) = \dfrac{1}{\alpha_0(j)}\left[\dfrac{1}{\lambda} + \sum_{k=1}^{j-1} \alpha_k(j) T(k)\right]$ if $\alpha_0(j) > 0$ and $j \in \{2, \ldots, K-1\}$.

Assumption A.2 implies that a trader, who faces a one-tick spread, submits a market order, thus $T(1) = \frac{1}{\lambda}$, that is, the average time between two arrivals. The expected waiting time of a *j*-limit order that never attracts a market order [i.e., such that $\alpha_0(j) = 0$] is obviously infinite. If $\alpha_0(j) > 0$, the expected waiting time of a *j*-limit order is a function of the expected waiting times of the subsequent orders that create a smaller spread. This implies that the expected waiting time function is recursive.

As the expected waiting time function is recursive, we can solve the game by *induction*. To see this, consider a trader who arrives when the spread is $s = 2$. The trader can submit either a market order or a one-tick limit order. The latter improves his execution price by one tick and results in an expected waiting time equal to $T(1) = 1/\lambda$. Solving Program (1) for patient and impatient traders, we determine $\alpha_k(2)$ for $k = 0$ and $k = 1$. If no trader submits a market order then $\alpha_0(2) = 0$, and the expected waiting time for a *j*-limit order is infinite for any $j \geq 2$ (Lemma 1). It follows that no spread larger than one tick can be observed in equilibrium. If instead we find that either patient or impatient traders submit market orders then $\alpha_0(2) > 0$ and we compute $T(2)$ using the Lemma 1. Next we proceed to $s = 3$ and so forth. As we proceed by induction, each type of trader has a unique optimal order placement strategy, and therefore the stationary equilibrium is unique.

The expected waiting time function has a recursive structure because our assumptions yield a simple ordering of the queue of unfilled limit orders in the book: a limit order is never executed before limit orders that create a smaller spread. Hence, the waiting time of a *j*-limit order is a function of the waiting times of limit orders that create smaller spreads. This ordering (and therefore Lemma 1) does *not* hold if buyers and sellers arrive randomly. Consider a buyer creating a spread of *j* ticks, followed by a seller creating a spread of $j'(j'<j)$ ticks. If the next trader is a second seller submitting a market order, the buyer is executed before the first seller. Assumption A.3 rules out this case. We consider the case in which this assumption does not hold in Section 5.

2.2 Equilibrium strategies

Recall that the payoff of a trader submitting a *j*-limit order is

$$\pi_i(j) \equiv j\Delta - \delta_i T(j),$$

and that this payoff is zero for a market order. Hence, a trader submits a *j*-limit order only if price improvement, $j\Delta$, exceeds waiting cost, $\delta_i T(j)$. A trader submitting a limit order expects to wait at least one period before execution. As the average duration of a period is $\frac{1}{\lambda}$, the expected waiting cost for a trader with type *i* is at least $\frac{\delta_i}{\lambda}$. It follows that the smallest spread trader *i* can establish is the smallest integer j_i^*, such that $\pi_i(j_i^*) = j_i^*\Delta - \frac{\delta_i}{\lambda} \geq 0$. We call j_i^* the "*reservation spread*" of a trader with type *i*. Let *CF*(*x*) denote the *ceiling function*, the smallest integer larger than or equal to *x* [e.g., *CF*(2.4) = 3, and *CF*(2) = 2]. Then the reservation spread is:

$$j_i^* \equiv CF\left(\frac{\delta_i}{\lambda\Delta}\right) \quad i \in \{P,I\}. \tag{2}$$

To exclude cases in which no trader submits limit orders, we assume that

$$j_P^* < K. \tag{3}$$

We call patient traders' reservation spread: *"the competitive spread"*, because no trader will post limit orders with smaller spreads. The reservation spread of a patient trader never exceeds that of an impatient one, but the two can be equal. We say that traders are *homogeneous* if patient and impatient traders have the same reservation spread: $j_P^* = j_I^* \overset{def}{=} j^*$. Otherwise we say that traders are *heterogeneous*.

2.2.1 The homogeneous case. If traders are homogeneous, then all traders prefer to submit a market order when the spread is less than or equal to j^* (by definition of the reservation spread). This implies that the waiting time for a j^*-limit order is one period with certainty. Hence,

$$\pi_i(j^*) \geq 0 \text{ for } i \in \{P,I\}. \tag{4}$$

Now, consider a trader who faces a book with an inside spread $s > j^*$. As $\pi_i(j^*) \geq 0$, the trader (patient or impatient) prefers to submit a j^*-limit order to a market order. Hence, a limit order which creates a spread larger than j^* is never executed. It follows that it is never optimal to post a spread larger than j^*. These observations lead us to Proposition 1.

Proposition 1. *When traders are homogeneous (i.e., $j_P^* = j_I^* = j^*$) then, in equilibrium, all traders submit a market order if $s \leq j^*$and submit a j^*-limit order if $s > j^*$.*

The equilibrium with homogeneous traders has two distinctive properties. First, all limit order traders post the competitive spread, j^*. Second, the spread oscillates between K and the competitive spread, and transactions take place only when the spread is competitive. Trade prices are either $A - j^*$ if the first trader is a buyer or $B + j^*$, if the first trader is a seller. We refer to this market as *strongly resilient*, since any deviation from the competitive spread is immediately corrected by the next trader.

The dynamics of the bid-ask spread in the homogeneous case look quite unusual, but they are not unrealistic. Biais, Hillion, and Spatt (1995) identify several typical patterns for the dynamics of the bid-ask spread in the Paris Bourse. Interestingly, they identify precisely the pattern we obtain when traders are homogeneous (Figure 3B, p. 1681): the spread alternates between a large and a small size, and all transactions take place when the spread is small. Given that this case requires that all traders have identical reservation spreads, we anticipate that this pattern is not frequent. It does, however, provide a useful benchmark for the results obtained in the heterogeneous trader case.

2.2.2 The heterogeneous case. Now consider *heterogeneous traders*, that is, $j_P^* < j_I^*$. In this case, there are spreads above patient traders' reservation spread for which impatient traders choose to submit market orders. We denote by $\langle j_1, j_2 \rangle$ the set: $\{j_1, j_1+1, j_1+2, \ldots, j_2\}$, that is, the set of all spreads between any two spreads $j_1 < j_2$ (inclusive).

Proposition 2. *Suppose traders are heterogeneous* $\left(j_P^* < j_I^*\right)$. *In equilibrium there exists a cutoff spread* $s_c \in \left\langle j_I^*, K \right\rangle$ *such that*:

1. *Facing a spread* $s \in \left\langle 1, j_P^* \right\rangle$*, both patient and impatient traders submit a market order.*
2. *Facing a spread* $s \in \left\langle j_P^* + 1, s_c \right\rangle$*, a patient trader submits a limit order and an impatient trader submits a market order.*
3. *Facing a spread* $s \in \langle s_c + 1, K \rangle$*, both patient and impatient traders submit limit orders creating a spread of* s_c.

The range of possible spreads is partitioned into three regions: (i) $s \leq j_P^*$, (ii) $j_P^* < s \leq s_c$, and (iii) $s > s_c$. The reservation spread of the patient trader, j_P^*, is the smallest spread observed in the market, while the cutoff spread s_c is the largest spread at which market orders are submitted. When the spread is larger than s_c, all traders submit limit orders. Limit orders that create a spread larger than s_c are never executed, hence are not submitted in equilibrium. Impatient traders demand liquidity by submitting market orders at spreads below s_c. Patient traders supply liquidity at spreads above their reservation spread and demand liquidity at spreads smaller than their reservation spread. It follows that the rate at which market orders are submitted is decreasing with the size of the spread.

In contrast to the homogeneous case, the inside spread may exceed an impatient trader's reservation spread, and yet this trader submits a market order (because in general $s_c > j_I^*$). The explanation for this result is as follows. Patient traders do not submit market orders when the inside spread is larger than j_P^*. Consequently, the expected time to execution of a j-limit order $\left(j > j_P^*\right)$ is strictly larger than one period. In particular $T^*\left(j_I^*\right) > \frac{1}{\lambda}$. For this reason, the expected waiting cost for an impatient trader posting his reservation spread exceeds, in general, his improvement in execution price [i.e., $\pi_I\left(j_I^*\right) = j_I^* \Delta - \delta_I T^*\left(j_I^*\right) < 0$]. In this case, the cutoff spread (s^c) at which impatient traders can profitably enter limit orders is strictly larger than their reservation spread. In many cases, the waiting costs increase so quickly with the size of the spread that impatient traders never find it optimal to enter limit orders at prices in the eligible range. In these cases, we set $s_c = K$. This occurs, for instance, when the cost of waiting for an impatient trader is sufficiently large.[10]

The cases in which $s_c < K$ and the case in which $s_c = K$ are qualitatively similar. The primary difference is that impatient traders never find it optimal to submit limit orders in the latter case while they do in the former case (when the spread is K). Henceforth, we focus our attention on the cases where $s_c = K$. This restriction has no significant impact on results, but shortens the presentation.

Proposition 3. *In equilibrium, there exist q spreads $(K \geq q \geq 2)$, $n_1 < n_2 < \ldots < n_q$, with $n_1 = j_P^*$, and $n_q = K$, such that the optimal order submission strategy is:*

- *An impatient trader submits a market order for any spread in $\langle 1, K \rangle$.*
- *A patient trader submits a market order when he faces a spread in $\langle 1, n_1 \rangle$, and submits an n_h-limit order when he faces a spread in $\langle n_h + 1, n_{h+1} \rangle$ for $h = 1, \ldots, q-1$.*

Thus, when a patient trader faces a spread n_{h+1} $(h \geq 1)$, he responds by submitting a limit order which improves the spread by $(n_{h+1} - n_h)$ ticks. This order establishes a new spread equal to n_h. This process continues until a market order arrives.

The next two propositions provide a closed-form solution for the expected waiting time function and for the equilibrium spreads. Let $\rho \equiv \frac{\theta_P}{\theta_I}$ be the ratio of the proportions of patient and impatient traders. Intuitively, when this ratio is smaller (larger) than 1, liquidity is consumed more (less) quickly than it is supplied since impatient traders submit market orders, and patient traders tend to submit limit orders. In equilibrium, the expected waiting time for a limit order is a function of ρ.

Proposition 4. *The expected waiting time function in equilibrium is:*

$$T^*(n_1) = \frac{1}{\lambda};\, T^*(n_h) = \frac{1}{\lambda}\left[1 + 2\sum_{k=1}^{h-1} \rho^k\right] \;\forall\, h = 2, \ldots, q-1; \tag{5}$$

and

$$T^*(j) = T^*(n_h) \quad \forall\, j \in \langle n_{h-1} + 1,\, n_h \rangle \;\forall\, h = 1, \ldots, q-1.^{11}$$

[10] Obviously $s_c = K$ if $j_I^* \geq K$. It is worth stressing that this condition is sufficient, but not necessary. In all the numerical examples below, j_I^* is much smaller than K, but $s_c = K$.

[11] We set $n_o = 0$ by convention.

Recall that a limit order is never executed before limit orders that create smaller spreads. For this reason, the choice of a spread is tantamount to the choice of a priority level in a waiting line: the smaller is the spread chosen by a trader, the higher is his priority in the queue of unfilled limit orders. Accordingly, the expected waiting time function increases with the spread. This property is consistent with evidence in Lo, McKinley, and Zhang (2001) who find that the time to execution of limit orders increases in the distance between the limit order price and the mid-quote.

Consider a trader facing a spread $n_{h+1}(h \leq q-1)$. In equilibrium, he submits an n_h-limit order (Proposition 3). He could reduce his expected time to execution by submitting an n_{h-1}-limit order, but chooses not to. Thus the following condition must hold:

$$n_h\Delta - T^*(n_h)\delta_P \geq n_{h-1}\Delta - T^*(n_{h-1})\delta_P, \quad \forall h \in \{2, \ldots, q-1\},$$

or

$$\Psi_h \equiv n_h - n_{h-1} \geq [T^*(n_h) - T^*(n_{h-1})]\frac{\delta_P}{\Delta}, \quad \forall h \in \{2, \ldots, q-1\}. \quad (6)$$

Now consider a trader facing a spread n_h. In equilibrium, this trader submits an n_{h-1}-limit order. Thus, he must prefer this limit order to a limit order which creates a spread of $(n_h - 1)$ ticks, which imposes

$$n_{h-1}\Delta - T^*(n_{h-1})\delta_P \geq (n_h - 1)\Delta - T^*(n_h - 1)\delta_P \quad \forall h \in \{2, \ldots, q\}.$$

In equilibrium, either (i) $T^*(n_h - 1) = T^*(n_h)$ or (ii) $T^*(n_h - 1) = T^*(n_{h-1})$. In both cases $T^*(n_h) > T^*(n_{h-1})$, hence

$$\Psi_h < [T^*(n_h) - T^*(n_{h-1})]\frac{\delta_P}{\Delta} + 1 \quad \forall h \in \{2, \ldots, q\}. \quad (7)$$

Combining Equations (6) and (7), we deduce that

$$\Psi_h = CF([T^*(n_h) - T^*(n_{h-1})]\frac{\delta_P}{\Delta}) = CF\left(2\rho^{h-1}\frac{\delta_P}{\lambda\Delta}\right), \forall h \in \{2, \ldots, q-1\}, \quad (8)$$

where the last equality follows from Proposition 4. Using Equation (8), we derive the set of equilibrium spreads $n_1, n_2, \ldots, n_q$.

Proposition 5*. The set of equilibrium spreads is given by:*

$$n_1 = j_P^*, \quad n_q = K,$$
$$n_h = n_1 + \sum_{k=2}^{h} \Psi_k \quad h = 2, \ldots, q-1,$$

where

$$\Psi_k = CF\left(2\rho^{k-1}\frac{\delta_P}{\lambda\Delta}\right),$$

and q is the smallest integer such that:

$$j_P^* + \sum_{k=2}^{q} \Psi_k \geq K. \tag{9}$$

We refer to $\Psi_h \overset{def}{=} n_h - n_{h-1}$ as *the spread improvement*, when the spread is n_h. The spread improvement is the number of ticks by which a limit order trader improves upon (undercuts) the best quotes. Thus, it is a measure of the aggressiveness of the submitted limit order: the larger is Ψ_h, the more aggressive is the limit order.

Proposition 5 presents the determinants of spread improvements. Spread improvements are larger when (i) the proportion of patient traders, θ_P, is large, (ii) the waiting cost, δ_P, is large, and (iii) the order arrival rate, λ, is small. In particular, when $2\delta_P\left(\frac{\theta_P}{\theta_I}\right)^{h-1} > \lambda\Delta$, a patient trader improves the spread by more than one tick, that is, $\Psi_h > 1$. Biais, Hillion, and Spatt (1995), among others, find that many limit orders in the Paris Bourse improve upon the prevailing bid-ask quotes by more than one tick.

A common mechanism is driving these results: a rise in expected waiting costs induces liquidity suppliers to bid more aggressively in order to shorten their execution time. Consider an increase in the proportion of patient traders. This increase immediately reduces the execution rate for limit orders as market orders become less frequent. Accordingly, the expected waiting time, $T(\cdot)$ and, thereby, the expected waiting cost for liquidity suppliers increase. Patient traders react by submitting more aggressive orders, thus Ψ_h increases for all $h > 1$. A similar reasoning applies when λ falls or δ_P rises.

2.2.3 Efficiency. What is the efficient outcome in our model? The price concession paid by liquidity demanders is earned by liquidity suppliers, *net* of their waiting costs. Hence, the bid-ask spread is just a transfer payment, and the waiting costs constitute a dead-weight loss. It follows that the total welfare—the sum of liquidity suppliers' and liquidity demanders' expected payoffs—is equal to the expected waiting costs borne by liquidity suppliers. We denote this expected cost by *EC*. An efficient outcome is such that traders use order placement strategies which result in the smallest possible value for *EC*.

Liquidity suppliers must wait *at least* one period before execution and their waiting cost, *per unit of time*, is *at least* δ_P; hence $EC \geq \frac{\delta_P}{\lambda}$. We deduce that a situation in which order placement strategies are such that $EC = \frac{\delta_P}{\lambda}$ is efficient. This can be attained under the following two conditions: (i) only the patient traders submit limit orders and (ii) all limit orders generate the competitive spread. In this case, *all* limit order traders (i) have the smallest possible waiting cost per unit of time (δ_P) and (ii) *always* wait exactly one period before execution. Thus $EC = \frac{\delta_P}{\lambda}$.

In general, the efficient outcome is not obtained *in equilibrium*. First, consider the homogeneous case. All the liquidity suppliers post the competitive spread, but some of them are impatient. As impatient traders bear relatively large waiting costs per unit of time $(\delta_I > \delta_P)$, the equilibrium outcome is inefficient (i.e., $EC > \frac{\delta_P}{\lambda}$). In the heterogeneous case, only patient traders provide liquidity; however, they strategically post spreads larger than the competitive spread. As a consequence, liquidity suppliers' average time to execution is strictly larger than one period and, for this reason, $EC > \frac{\delta_P}{\lambda}$. In summary, our model uncovers two possible sources of inefficiency in a limit order market: (i) impatient traders sometimes submit limit orders and (ii) patient traders post spreads larger than the competitive spread. Both features induce excessive waiting costs compared to the efficient outcome.

2.3 The determinants of market resiliency

Suppose that a liquidity shock (a succession of market orders in our model) causes the spread to increase. In a resilient market, it takes a small number of quote updates for the spread to revert to its former level. Hence, market resiliency can be measured by the average number of orders observed before the spread reverts to its competitive level. The smaller is this number, the larger is market resiliency. Degryse *et al.* (2003), for instance, use this method. It is difficult to derive analytically a general expression for this *order-based* measure of market resiliency in our model. Instead, we use another closely related measure. We measure market resiliency as the *probability* that the spread reverts to its competitive level before the next transaction occurs. The larger is this probability, the larger is market resiliency. We have checked numerically that all our implications regarding our measure of market resiliency (for instance Corollary 1) hold for the order-based measure of resiliency as well.[12]

Suppose that initially the spread is at its competitive level, j_P^*, and that a succession of market orders enlarges the spread to its maximal level of K. Let R be the probability that the spread reverts to the competitive level, j_P^*, before the next transaction. We will take R as being our measure of market resiliency. When traders are homogeneous, a deviation from the

[12] Our measure of market resiliency is based on the *inside spread*. In reality, the speed at which *depth* at given quotes recovers after a liquidity shock is another dimension of market resiliency [see Coppejans, Domowitz, and Madhavan (2003) and Degryse *et al.* (2003)]. We cannot study this dimension, since all orders are of the same size in our model, and we do not allow orders to queue at the same price.

competitive spread is immediately corrected and $R = 1$. When traders are heterogeneous, a streak of $q - 1$ consecutive patient traders is required to narrow the spread to the competitive level conditional on a current spread of K ticks (see Proposition 3). Thus $R = \theta_P^{q-1} < 1$. Notice that q is endogenous and is a function of all the exogenous parameters [see Equation (9)]. Consequently, market resiliency is determined jointly by the proportion of patient traders, the order arrival rate, trader's waiting costs, and the tick size.

Corollary 1. *When traders are heterogeneous, the resiliency of the limit order book, R, increases in the proportion of patient traders, θ_P, and the waiting cost, δ_P, but decreases in the order arrival rate, λ.*

The factors which enlarge (lower) spread improvements have a positive (negative) effect on the resiliency of the limit order book. As the proportion of patient traders rises, or as waiting costs increase, limit order traders improve upon the inside spread by larger amounts, thereby increasing the market resiliency. With an increase in the order arrival rate, limit order traders become less aggressive in their price improvements. More orders are required to bring the spread to the competitive level, and thus resiliency declines.[13] We postpone the analysis of the effect of the tick size on market resiliency to Section 4.

2.4 Numerical examples

We study three examples representing three cases of interest: (a) traders are homogeneous, (b) traders are heterogeneous and patient traders dominate the trading population $\left(\rho = \frac{\theta_P}{\theta_I} \geq 1\right)$, and (c) traders are heterogeneous and impatient traders dominate the trading population ($\rho < 1$). Our goal here is to show that the dynamics of the limit order book are strikingly different in these three cases.

In each example, the tick size is $\Delta = \$\frac{1}{16}$ and the arrival rate is $\lambda = 1$. The lower and upper price bounds of the limit order book are $B\Delta = \$20$ and $A\Delta = \$21.25$. The maximal spread is $K = 20$ ($K\Delta = \$1.25$). The remaining parameters are presented in Table 1.

In Example 1, the waiting costs are such that traders are homogeneous $\left(j_P^* = j_I^* = 2\right)$. In Examples 2 and 3, the waiting costs are such that traders are heterogeneous $\left(j_P^* = 1 \text{and } j_I^* = 3\right)$. The composition of the trading population is different in Examples 2 and 3: patient traders dominate in Example 2 ($\rho = 1.22$), whereas impatient traders dominate in Example 3 ($\rho = 0.81$). In Example 1, the competitive spread is two ticks. In Examples 2 and 3, the competitive spread is one tick.

[13] The average time elapsed until the spread reverts to its competitive level could also be used to measure resiliency. However, the effect of increasing the order arrival rate on this time-based measure of market resiliency is ambiguous. While it increases the number of quote updates before the spread reverts to its former level, at the same time it reduces the average time between orders.

Table 1
Parameter values for the three examples

	Example 1	Example 2	Example 3
δ_P	0.075	0.050	0.050
δ_I	0.100	0.125	0.125
θ_P	Any value	0.55	0.45

Table 2
Equilibrium order placement strategies

	Example 1		Example 2		Example 3	
Current spread	Type P	Type I	Type P	Type I	Type P	Type I
1	0	0	0	0	0	0
2	0	0	1	0	1	0
3	2	2	1	0	1	0
4	2	2	3	0	3	0
5	2	2	3	0	3	0
6	2	2	3	0	5	0
7	2	2	6	0	6	0
8	2	2	6	0	7	0
9	2	2	6	0	8	0
10	2	2	9	0	9	0
11	2	2	9	0	10	0
12	2	2	9	0	11	0
13	2	2	9	0	12	0
14	2	2	13	0	13	0
15	2	2	13	0	14	0
16	2	2	13	0	15	0
17	2	2	13	0	16	0
18	2	2	13	0	17	0
19	2	2	18	0	18	0
20	2	2	18	0	19	0

2.4.1 Order placement strategies. Table 2 reports the equilibrium strategies for patient and impatient traders (Types P and I). These strategies derive from Proposition 5. Each row corresponds to a spread when a trader arrives. Each entry gives the trader's optimal order for spreads on and off the equilibrium path. The limit orders are expressed in ticks, and 0 indicates the placement of a market order. For instance, in Example 2, when the inside spread is eight ticks, a patient trader submits a limit order which creates a spread of six ticks, and an impatient trader submits a market order.

In Example 1 (homogeneous case), patient and impatient traders submit a limit order creating a spread of two ticks when the spread is larger than their reservation spread. Otherwise they submit a market order. Thus, the inside spread oscillates between the maximal spread of 20 ticks and the competitive spread $\left(j_P^* = \text{two ticks}\right)$. In Examples 2 and 3 (heterogeneous case), patient traders supply liquidity, and impatient traders demand liquidity. In contrast to Example 1, limit order traders do not necessarily post the competitive spread $\left(j_P^* = 1\right)$. In fact, in Example 2, the spreads on the equilibrium path are: {1, 3, 6, 9, 13, 18, 20}; while in Example 3 these are: {1, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 17, 18, 19, 20}. Traders place more aggressive limit orders when patient traders dominate the trading population (Example 2). Actually, spread improvements are larger than one tick for all spreads on the equilibrium path in Example 2. In contrast, in Example 3, spread improvements are equal to one tick in most cases.

2.4.2 Expected waiting time. Figure 1 presents the expected waiting times of limit orders in Examples 2 and 3 as functions of the spread. In each example, the expected waiting time increases with the spreads on the equilibrium path and remains constant over spreads that are not reached in equilibrium. The expected waiting times are uniformly smaller in Example 3 than in Example 2. This explains the differences in the optimal strategies. When the proportion of patient traders is small, as in Example 3, patient traders are not aggressive, *because* they expect a fast execution.

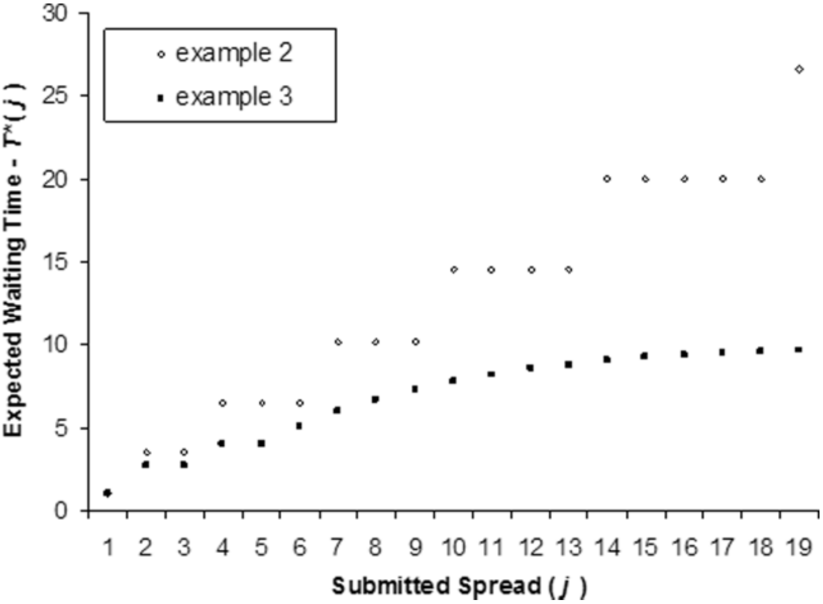

Figure 1
Expected waiting time
The figure reports the expected waiting times of limit orders as a function of the spread they create in Examples 2 and 3. Expected waiting times for Example 2 are depicted by White Diamonds and for Example 3 by Black Squares.

2.4.3 Book dynamics and resiliency. Figure 2 reports the evolution of the limit order book for a sequence of 40 trader arrivals in Examples 2 and 3. In each period, it gives the state of the limit order book *after* the order submission of the trader arriving in this period. Figure 3 depicts the corresponding dynamics of the inside spread and its mid-point (black dots). Initially the spread is equal to $K = 20$ ticks. This is the state of the book after the arrival of several market orders. How fast does the spread revert to the competitive level (one tick)?

Example 2 - A Resilient Book (ρ = 1.222)

Period	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Trader	BI	SP	BP	SI	BI	SI	BP	SP	BI	SP	BI	SP	BP	SP	BI	SP	BP	SP	BP	SI	BP	SP	BP	SP	BI	SI	BP	SI	BP	SP	BP	SI	BI	SP	BI	SI	BI	SP	BP	SI
21 1/4	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
21 3/16																																								
21 1/8		o	o	o																																				
21 1/16																																								
21																																								
20 15/16								o		o		o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o		o	o	o
20 7/8																																								
20 13/16																																								
20 3/4														o		o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o		o						
20 11/16																																								
20 5/8																		o				o		o						o										
20 9/16																	b	b	b		b	b	b	b	b		b		b	b	b									
20 1/2																																								
20 7/16																																								
20 3/8													b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b				b	
20 5/16			b																																					
20 1/4																																								
20 3/16																																								
20 1/8							b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b
20 1/16																																								
20	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b

Example 3 - A Weakly Resilient Book (ρ = 0.818)

Period	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Trader	BI	SP	BP	SI	BI	SI	BP	SP	BI	SP	BI	SP	BP	SP	BI	SP	BP	SP	BP	SI	BP	SP	BP	SP	BI	SI	BP	SI	BP	SP	BP	SI	BI	SP	BI	SI	BI	SP	BP	SI
21 1/4	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
21 3/16		o	o	o				o		o		o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
21 1/8														o		o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
21 1/16																		o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
21																						o	o	o	o	o	o	o	o	o	o	o	o	o	o	o		o	o	o
20 15/16																								o						o	o	o		o						
20 7/8																																								
20 13/16																																								
20 3/4																																								
20 11/16																																								
20 5/8																																								
20 9/16																																								
20 1/2																																								
20 7/16																																								
20 3/8																															b									
20 5/16																							b	b	b		b		b	b	b	b	b	b	b				b	
20 1/4																			b		b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b
20 3/16																	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b
20 1/8													b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b
20 1/16			b				b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b
20	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b

Figure 2
Book simulation
The figure presents the evolution of the limit order book in Examples 2 and 3 for a given sequence of 40 trader arrivals. For each period, the figure indicates the type of the trader arriving in this period. "BP" and "BI" indicate the arrival of a "Patient Buyer" and an "Impatient Buyer" respectively. "SP" and "SI" indicate the arrival of a "Patient Seller" and an "Impatient Seller" respectively. For each period, the figure gives the state of the book after the order submission by the trader arriving in this period. Letter "b" (respectively 'o') at a given price indicates the presence of a buy (respectively sell) limit order at this price.

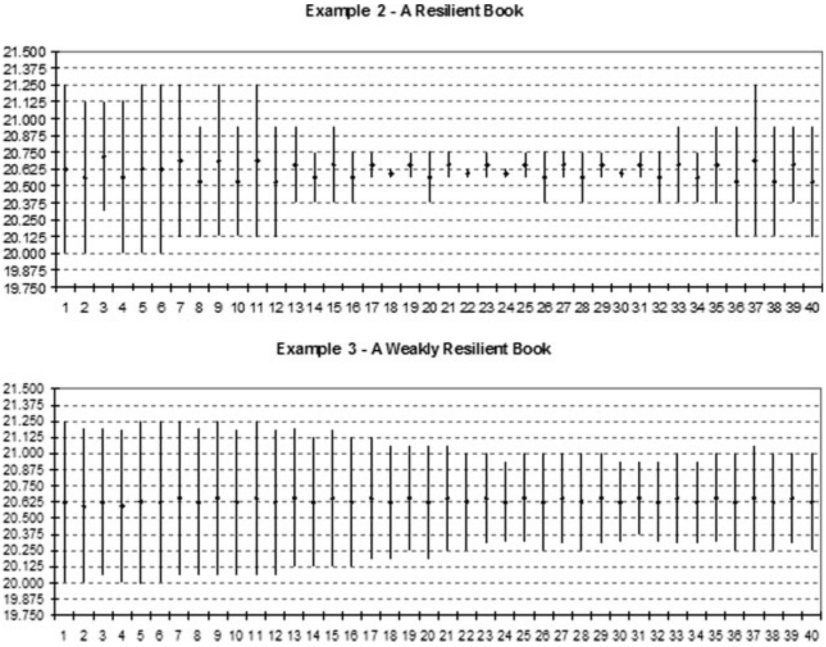

Figure 3
Spread evolution
The figure presents the evolution of the size of the inside spread (vertical line) and the mid-quote (dots) in Examples 2 and 3 for the sequence of trader arrivals depicted in Figure 2.

It is apparent from Figure 3 that the competitive spread is reached much more quickly in Example 2 than in Example 3. In fact, in Example 3, the quoted spread remains much larger than the competitive spread during all 40 periods depicted in Figures 2 and 3. In contrast, in Example 2, the inside spread reaches the competitive level for the first time after 18 periods and then it remains close to this level. As traders' types in each period are identical in each example, this is due to the fact that in Example 2, patient traders use more aggressive limit orders in order to speed up execution.[14] This bidding behavior explains why the market appears much more resilient in Example 2 than in Example 3. Accordingly, our measure indicates that the resiliency of the market is much larger in Example 2, $R = 0.55^6 \simeq 0.02$, than in Example 3, where $R = 0.45^{17} \simeq 1.27 \times 10^{-6}$.

[14] If realizations for traders' types were not held constant, an additional force would make small spreads more frequent when $\rho \geq 1$. In this case, the liquidity offered by the book is *consumed less rapidly*, since the likelihood of a market order is smaller than when $\rho < 1$. Thus, the inside spread has more time to narrow between market order arrivals.

Summary. When traders are homogeneous, any deviation from the competitive spread is immediately corrected. This is not the case in general when traders are heterogeneous. In the latter case, the market is more resilient when $\rho \geq 1$ than when $\rho < 1$. Thus, although the equilibrium of the limit order market is unique, three patterns for the dynamics of the spread emerge: (a) *strongly resilient*, when traders are homogeneous, (b) *resilient*, when traders are heterogeneous and patient traders dominate ($\rho \geq 1$), and (c) *weakly resilient*, when traders are heterogeneous and impatient traders dominate ($\rho < 1$). Market resiliency clearly depends on the composition of the traders' population.

2.5 Distribution of spreads

In this section, we derive the equilibrium probability distribution of the bid-ask spread when traders are heterogeneous (when they are homogeneous, all transactions take place at the competitive spread). We show how the distribution of spreads depends on the composition of the trading population.

Proposition 3 shows that the equilibrium spread takes q different values, $n_1 < n_2 < \ldots < n_q$. A patient trader submits a n_{h-1}-limit order when the spread is n_h ($h = 2, \ldots, q$) and a market order when the spread is n_1. An impatient trader always submits a market order. These strategies yield the following dynamics for the bid-ask spread. If the spread is n_h $(h = 2, \ldots, q-1)$, there is a probability θ_P that it becomes n_{h-1} and a probability θ_I that it becomes n_{h+1} when a new trader arrives in the market. If the spread is n_1, the inside spread becomes n_2 when a new trader arrives in the market. If the spread is K, it remains unchanged with probability θ_I or it becomes n_{q-1} with probability θ_P, when a new trader arrives in the market. Hence, the random sequence of observed spreads is a finite Markov chain with $q \geq 2$ states. The transition matrix of this Markov chain is:

$$W = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ \theta_P & 0 & \theta_I & \cdots & 0 & 0 \\ 0 & \theta_P & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \theta_I \\ 0 & 0 & 0 & \cdots & \theta_P & \theta_I \end{pmatrix}$$

The j^{th} entry in the h^{th} row of this matrix is the probability that the size of the spread changes to n_j conditional on the spread n_h $(j, h = 1, \ldots, q)$. We denote the stationary probabilities of this Markov chain by $u_1, \ldots, u_q$, where u_h is the stationary probability of a spread of size n_h.

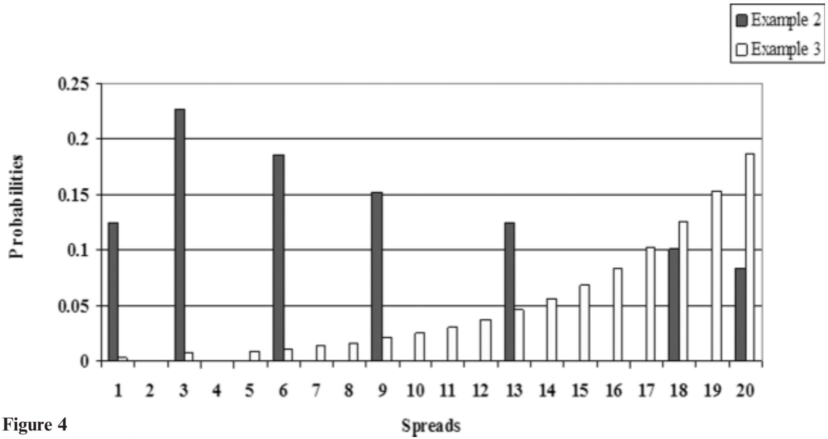

Figure 4
Spread distribution
The figure presents the stationary probability distribution of the inside spread in Examples 2 (grey bars) and 3 (white bars). For each inside spread on the equilibrium path, the figure gives the probability of occurrence of this inside spread.

Lemma 2. *The spread has a unique stationary probability distribution given by:*

$$u_1 = \frac{\theta_P^{q-1}}{\theta_P^{q-1} + \sum_{i=2}^{q} \theta_P^{q-i} {\theta_I}^{i-2}}, \textit{ and } u_h = \frac{\theta_P^{q-h}\theta_I^{h-2}}{\theta_P^{q-1} + \sum_{i=2}^{q} \theta_P^{q-i} {\theta_I}^{i-2}} \\ h = 2, \ldots, q. \tag{10}$$

Observe that for $h, h' \in \{2, 3, \ldots, q\}$ with $h > h'$, Lemma 2 implies that

$$\frac{u_h}{u_{h'}} = \rho^{h'-h} \textit{ and } \frac{u_h}{u_1} = \frac{1}{\rho^{h-1}\theta_I}. \tag{11}$$

This remark yields the following proposition.

Proposition 6. *For a given tick size and waiting costs*:

1. *If* $\rho<1$, $u_h>u_{h'}$ *for* $1 \leq h'<h \leq q$. *Thus, the distribution of spreads is skewed toward higher spreads in weakly resilient markets* (*i.e.*, $\rho < 1$).
2. *If* $\rho>1$, $u_h \leq u_{h'}$ *for* $2 \leq h'<h \leq q$.[15] *Thus, the distribution of spreads is skewed toward lower spreads in resilient markets* (*i.e.*, $\rho > 1$).

Hence, small spreads are more frequent when $\rho > 1$ than when $\rho < 1$. This reflects the fact that markets dominated by patient traders are more resilient than markets dominated by impatient traders. Figure 4 depicts

[15] The inequality, $u_h<u_{h'}$, does not necessarily hold for $h' = 1$, when $\rho>1$. Actually, the smallest inside spread can only be reached from higher spreads, while other spreads can be reached from both directions ($n_q = K$ can be reached either from n_{q-1} or from n_q itself). This implies that the probability of observing the smallest possible spread is relatively small for all values of ρ.

the stationary distribution in Examples 2 and 3. Clearly, the distribution of spreads is skewed toward high spreads in Example 3 where $\rho < 1$ and toward low spreads in Example 2 where $\rho > 1$.

The expected dollar spread is given by:[16]

$$ES^m = \sum_{h=1}^{q} u_h n_h^m. \tag{12}$$

Using this expression, the expected spread is \$0.525 in Example 2 and \$1 in Example 3. The average spread is smaller in the more resilient market, because small spreads are more frequent in this case.

3. Implications for Empirical Analysis

In this section, we discuss the implications of our model for empirical studies of limit order markets. We restrict attention to the heterogeneous case. We study the relationships between measures of trading activity (the order arrival rate and the time between trades) and market resiliency. Then we analyze the impact of the state of the book on order aggressiveness. Finally, we examine the implications of changes in the proportion of patient traders on intraday patterns and discuss the liquidity effects of changes in ownership.

3.1 Trading activity, spreads and market resiliency

3.1.1 Fast versus slow markets. We compare market resiliency and the expected spread across two markets "*Fast*" (F) and "*Slow*" (S), which differ only in the order-arrival rates, with $\lambda_F > \lambda_S$.[17] Thus, the average waiting time between orders in market F is less than in market S. We denote by $n_h(\lambda_k)$ the h^{th} smallest spread in the set of spreads on the equilibrium path in market $k \in \{F, S\}$ and by q_k the number of spreads in this set. Using Proposition 5 and Corollary 1, we obtain the following result.

Corollary 2. *In two markets F and S with order-arrival rates,* $\lambda_F > \lambda_S$,

1. *The equilibrium spreads satisfy:*

   (a) $n_h\,(\lambda_F) \le n_h\,(\lambda_S)$, *for* $h \le q_S$ *and*
   (b) $n_h\,(\lambda_F) \le n_{qs}\,(\lambda_S)$, *for* $q_S \le h \le q_F$.

2. *The Slow market is more resilient than the Fast market.*

[16] Recall that a superscript "m" indicates variables expressed in monetary terms, rather than in number of ticks (i.e., $n_h^m = n_h\Delta$).

[17] In our model, equilibrium spreads are determined by the ratio $\frac{\delta_P}{\lambda}$ (see Proposition 5). For this reason, the results for an increase in the arrival rate translate immediately to results for a decline in the waiting costs δ_P.

Table 3
Expected dollar spreads and order arrival rates

			θ_{P}		
λ	0.45	0.475	0.5	0.525	0.55
1	1	0.9	0.71	0.57	0.525
4/5	1.01	0.91	0.73	0.67	0.6
2/3	1.01	0.86	0.73	0.65	0.59
1/2	1.02	0.85	0.78	0.68	0.67
1/3	0.97	0.84	0.84	0.78	0.78
1/5	0.99	0.91	0.91	0.89	0.90

The first part of the corollary means that the support of possible spreads in the Fast market is shifted to the left compared to the support of possible spreads in the Slow market. To understand this result, compare two limit order traders, one in the Fast market and the other in the Slow market, with equal priority in the queues of limit orders. The expected waiting time of the trader in the Fast market is smaller. Thus, limit order traders in the Fast market require less compensation for taking a given position in the queue of limit orders. The result is that spreads tend to be narrower in the Fast market (first part of the proposition). On the other hand, spread improvements are larger, and the spread narrows more quickly after liquidity shocks in the Slow market (see the discussion following Proposition 5). Hence, the Slow market is more resilient.

These two effects have an opposite impact on the average spread. We cannot determine analytically which market, Fast or Slow, has the smallest spread. Computations show that increasing the order arrival rate reduces the expected spread for a wide range of parameters' values, which suggests that the first effect dominates in many cases. As an illustration of this claim, Table 3 reports the equilibrium expected dollar spread for various pairs (θ_P, λ). The values of the other parameters are as in Example 3 (i.e., $\Delta = \frac{1}{16}$, $K^m = 1.25$, $\delta_P = 0.05$, $\delta_I = 0.125$).[18] The correlation between the average spread and the order arrival rate is negative and equal to –0.37 in the sample defined by the table. This suggests that the average spread declines as the order arrival rate increases.

In summary, the model generates two predictions. First, the average spread tends to be inversely related to the order arrival rate. More surprising, maybe, the model implies a negative relationship between market resiliency and the order arrival rate.

[18] The condition $s_c = K$ holds for all parameter values considered in this table. Hence, we use Proposition 5, Lemma 2, and Equation (12) to compute the equilibrium expected spreads.

We interpret λ as the *long-run* arrival rate of orders, including both market *and* limit orders. It is a measure of trading activity in a given stock. Demsetz (1968) suggests that the primary determinant of the long-run order arrival rate is the number of shareholders. We conjecture that another related factor is the public float of the stock. As these variables are not affected by variations in the state of the book, it is natural to consider λ as independent from the state of the limit order book as well, just as it is exogenous in our model.[19]

We recognize that, in reality, the intraday arrival rate deviates from the long-run arrival rate, and that these deviations may in part be explained by transient changes in the state of the book. We are not aware of empirical evidence from limit order markets pointing out in this direction. At any rate, if such a relation exists, it is not captured by our model.

3.1.2 Durations between trades and market resiliency.

The order arrival rate is one measure of market activity. Market activity can also be measured by the average time between trades. There has been a considerable interest recently in modeling the time between trades. For instance, the *Autoregressive Conditional Duration* approach, pioneered by Engle and Russell (1998), postulates the expected duration between trades as a function of pre-determined variables including past realizations of the duration between trades. There is very little theory that endogenizes the time between trades in limit order markets.[20] This scarcity makes it difficult to specify and interpret conditional duration models.

In our model, the time between trades is endogenous, and it depends on the size of the spread. Let D_h denote the expected time elapsing between two consecutive transactions, *conditional* on the first transaction taking place when the spread is n_h. We call this variable a *conditional duration*.

Corollary 3. *In equilibrium, the conditional duration is:*

$$D_h = \frac{1-\theta_P^{h+1}}{\lambda\theta_I} \text{ for } 1 \leq h < q; \text{ and } D_q = \frac{1-\theta_P^{q}}{\lambda\theta_I}. \tag{13}$$

Hence, the conditional duration (i) increases with the size of the inside spread, (ii) decreases with the order arrival rate, and (iii) increases in the proportion of patient traders.

The conditional duration decreases as the order arrival rate rises, *other things being equal*, and is positively related to the spread: the larger is the

[19] The rates at which market and limit orders arrive are endogenous since traders optimally choose between these two types of orders. As expected, they depend on the spread (see Propositions 1 and 2). This is the *overall* order arrival rate which is exogenous in our model.

[20] Easley and O'Hara (1992) endogenize the time between transactions in a dealer market. In their model, the timing of trades is driven by the existence of new information.

spread at which a trade occurs, the larger is the average time until the next transaction. Notice that there is an interesting contrast with the findings obtained in Easley and O'Hara (1992). In their model, the spread depends on the time elapsed since the last transaction while in our model, the expected time until the next transaction depends on the size of the inside spread. This suggests that the spread should be used as an explanatory variable in empirical models of conditional duration. Finally, the conditional duration increases with the proportion of patient traders, θ_P. As θ_P rises, the probability that a trader submits a market order declines which delays the next transaction.

In our model, conditional duration and market resiliency are governed by the same factors. These factors include the order arrival rate or the proportion of patient traders. Consider an increase in the proportion of patient traders. Other things being equal, it leads to (i) a larger conditional duration (Corollary 3) and simultaneously to (ii) greater market resiliency (Corollary 1). Similarly an increase in the order arrival rate results in (i) a smaller conditional duration (Corollary 3) and in (ii) lower market resiliency (Corollary 1). Consequently, our model predicts a positive association between the average time between trades, conditional on the size of the spread, and market resiliency.

3.2 State of the book and order aggressiveness

The amount by which limit order traders improves upon prevailing quotes ("the spread improvement") is a measure of order aggressiveness.[21] Proposition 5 has the following implication.

Corollary 4. *Spread improvements depend on the size of the inside spread. Spread improvements increase with the size of the spread when* $\rho > 1$ *and decrease with the size of the spread when* $\rho < 1$.

As an illustration, compare the bidding strategies in Examples 2 and 3 in Table 2. When they face a spread of six ticks, limit order traders undercut the posted spread by three ticks in Example 2 ($\rho > 1$) and only one tick in Example 3 ($\rho < 1$). When they face a spread of three ticks, limit order traders undercut by two ticks in Examples 2 and 3. Hence, their aggressiveness when they face a spread of three ticks is (a) smaller (than when they face a spread of six ticks) in Example 2 but (b) larger in Example 3.

The explanation for this result is as follows. Consider a patient trader who faces a spread n_h. The spread improvement chosen by the trader results from the following trade-off. The cost of a large spread improvement is that it results in a worse execution price for the trader. The benefit

[21] Several empirical papers have analyzed the relationship between the state of the book and the aggressiveness of incoming orders. See, for instance, Griffith *et al.* (2000), and Biais, Hillion, and Spatt (1995).

is that it results in a smaller waiting cost. Hence, the trader's decision hinges upon a comparison between the size of the spread improvement and the resulting reduction in waiting cost. In equilibrium, the optimal spread improvement is equal to the expected reduction in the waiting cost, rounded up to the nearest integer. This follows from Equation (8):

$$\Psi_h = CF\left([T^*(n_h) - T^*(n_{h-1})]\frac{\delta_P}{\Delta} \right).$$

When $\rho > 1$, the difference $T^*(n_h) - T^*(n_{h-1})$ increases in n_h (see Proposition 4). It follows that liquidity suppliers are willing to offer larger spread improvements when the spread is large. In contrast, when $\rho < 1$, the difference $T^*(n_h) - T^*(n_{h-1})$ decreases in n_h; thus, liquidity suppliers are willing to make larger spread improvements at small spreads.

For stocks listed on the NYSE, Engle and Patton (2003) find that the change in the log of the best ask (bid) price is negatively (positively) related to the size of the spread. This means that the amount by which traders improve upon prevailing quotes is related to the size of the spread, as predicted by Corollary 4.[22] This corollary also implies that the direction of this relationship is affected by the proportion of patient traders. This is an additional prediction, which could be tested in future empirical work.

3.3 Changes in the mix of patient versus impatient traders

In our model, market resiliency, spread improvements and the distribution of spreads are functions of the proportion of patient traders. Unfortunately, the proportion of patient traders cannot be directly observed. Hence, it is hard to directly test the model predictions on the effects of a change in the proportion of patient traders.

One approach is to use a proxy for the proportion of patient traders. The model suggests that the proportion of limit orders in the flow of market and limit orders as a proxy. In equilibrium, patient traders tend to submit limit orders whereas impatient traders tend to submit market orders. In fact, conditional on a spread above the competitive spread, the probability of observing a limit order is θ_P, the proportion of patient traders. Another approach is to exploit predictable changes in the proportion of patient traders. We describe this approach in the following subsections.

3.3.1 Intraday variations in the proportion of patient traders and intraday patterns. Traders' impatience is likely to increase toward the end of the trading day. The inability to trade overnight is a binding constraint for many investors which makes them eager to trade as the end of the day

[22] A caveat is in order here since the NYSE is a hybrid market. There is a possibility that Engle and Patton's findings are driven by the actions of NYSE specialists rather than those of limit order traders. We are not aware of evidence on this issue from pure order-driven markets.

approaches.[23] Furthermore, many institutions mark their positions to market at the end of the day; thus, they prefer to trade closer to that deadline. Also, some traders act as implicit market markers in limit order markets and may be keen to unload their inventory before the end of the day to avoid an exposure to the overnight risk.

For these reasons, we conjecture that the proportion of impatient traders is larger at the end of the trading day than in earlier periods. Under this conjecture, comparing measures of market liquidity in the last period of the day (say the last half-hour) with these measures in an earlier period is *like* analyzing the impact of an *increase* in the proportion of impatient traders in our model (holding other parameters fixed). Several testable implications follow from this remark:

- Limit order traders submit less aggressive orders at the end of the day than in an earlier period. This prediction derives from Proposition 5 which establishes that spread improvements are small when the proportion of impatient traders is large.
- As a result, market resiliency is smaller at the end of the day than in an earlier period (see Corollary 1 and the discussion following the corollary).
- The spread is larger at the end of the day than in an earlier period (see Proposition 6 and the discussion in Section 2.4).
- The conditional durations between trades are smaller at the end of the day than during the day (see Corollary 3).

Testing these predictions empirically is not entirely straightforward. The information asymmetry is high at the beginning of the trading day and thus is likely to influence measures of market liquidity at this time. For instance, Madhavan, Richardson, and Roomans (1997) show empirically that the adverse selection component of the spread is large at the beginning of the day and declines thereafter. Hence, in order to avoid confounding effects due to asymmetric information, our predictions should be tested by comparing measures of market liquidity (limit order aggressiveness, resiliency, spreads, and conditional durations) in the last period of the trading day and in an earlier period, which is not too close to the opening of the trading session.

Some of our implications are consistent with stylized facts. It is well known that spreads and trading frequency decline from the opening of the trading day on but peak again at the end of trading sessions. For instance, Biais, Hillion, and Spatt (1995) and Chung *et al.* (1999) observe this pattern for the Paris Bourse and the NYSE, respectively. Although the quotes in the NYSE are in part determined by the specialist, Chung

[23] Recent experimental findings by Bloomfield, O'Hara, and Saar (2002) show that when liquidity traders are assigned a trading target, they switch from limit to market orders at the end of trading sessions.

et al. (1999) find that the increase in the spread is driven by the order placement decisions of limit order traders. The *joint* peak in spreads, and trade rates at the end of the trading day has proved difficult to explain in asymmetric information models. Actually, as pointed out by Foster and Viswanathan (1993), these models predict an inverse relationship between spreads and trading activity. Our model offers a complementary explanation. At the end of the day, the proportion of impatient traders increases. This translates into a decline in the conditional duration and thereby more frequent trades. Patient traders exploit this impatience by bidding less aggressively which results in larger spreads.

Pagano and Schwartz (2003) analyze the impact of the introduction of a closing call auction in the Paris Bourse. This closing auction offers another trading opportunity and should decrease traders' impatience toward the end of the day. Our model predicts that this should lead to a decline in the spread in the last half-hour of the trading day. This is precisely what is observed by Pagano and Schwartz (2003).[24] Notice that traditional theory would rather predict the opposite, as the closing auction is likely to draw liquidity away from the continuous market. Finally, Tkach and Kandel (2004) show that the time to execution of limit orders in Tel Aviv Stock Exchange declines toward the end of the trading day, in line with our predictions.

3.3.2 Liquidity effects of changes in ownership structures. We expect cross-sectional variations in the mix of patient and impatient traders to be related to variations in institutional ownership. Stocks that are predominantly owned by index funds should feature a larger proportion of impatient traders, since their managers must trade rapidly to minimize their fund's tracking error. This line of reasoning suggests to test the model by analyzing long-run liquidity effects on stocks that are added to (removed from) a widely followed index. Beneish and Whaley (1996), among others, document a substantial increase in the proportion of index funds owning a stock when it becomes listed in such an index. Our model predicts that the resulting increase in the proportion of impatient traders should manifest itself in a decline in market resiliency (Corollary 1) and smaller spread improvements (Proposition 5). Naturally, an inclusion in the index may have other effects on the stocks as well.

4. Tick Size and Market Resiliency

The tick size, that is, the minimal price variation, has been reduced in many markets in recent years. The rationale for this move was to reduce the trading costs of investors. In this section, we examine the effect of a

[24] "Kaniel and Liu (2000) can also explain this result."

change in the tick size, and show that a reduction in the tick size impairs market resiliency, and may have adverse effects on the spread.[25]

We assume that a change in the tick size does not affect the monetary values of the boundaries: $A^m = A\Delta$ and $B^m = B\Delta$, which implies that K^m is fixed as well.

To better convey the intuition, it is useful to consider the polar case in which there is no minimum price variation, that is, $\Delta = 0$. In this case, prices and spreads can be expressed solely in monetary terms; in what follows, we index all spreads by a superscript "*m*" to indicate that they are expressed in dollars. When the tick size is zero, a trader's reservation spread is exactly equal to his expected waiting cost until the arrival of the next trader, i.e., $j_i^{*m} = \frac{\delta_i}{\lambda}$ $(i \in \{P, I\})$. We denote by $T(j^m)$ the expected waiting time for a limit order trader who creates a spread of j^m dollars. Let

$$\rho^c \overset{def}{=} \frac{K^m\lambda - \delta_P}{K^m\lambda + \delta_P}. \tag{14}$$

Notice that $0 < \rho^c < 1$, since $j_P^{*m} < K^m$ by assumption [Equation (3)]. The next proposition extends Propositions 4 and 5 to the case in which there is no mandatory minimum price variation, but $\rho > \rho^c$.[26]

Proposition 7. *Suppose that $\Delta = 0$. If $\rho > \rho^c$ and $\delta_P > 0$, the equilibrium is as follows*:

1. *The impatient traders never submit a limit order.*
2. *There exist q_0 spreads $n_1^m < n_2^m < \ldots n_{q_0}^m$, with $n_1^m = \frac{\delta_P}{\lambda}$ and $n_{q_0}^m = K^m$ such that a patient trader submits an n_h^m-limit order when he faces a spread in $\left(n_h^m, n_{h+1}^m\right]$ and a market order when he faces a spread smaller than or equal to n_1^m.*[27]
3. *The spreads are: $n_h^m = n_{h-1}^m + \Psi_h^m(0)$, where $\Psi_h^m(0) = \left(2\rho^{h-1}\right)\frac{\delta_P}{\lambda}$, for $h = 2, \ldots q_0 - 1$ and the stationary probability of the h^{th} spread is u_h, as given in Section 2.5.*
4. *The expected waiting time function is as follows: $T^*\left(n_1^m\right) = \frac{1}{\lambda}$; $T^*\left(n_h^m\right) = \frac{1}{\lambda}\left[1 + 2\sum_{k=1}^{h-1}\rho^k\right]$ $\forall h \in \{2, \ldots, q_0 - 1\}$; and $T^*\left(j^m\right) = T^*\left(n_h^m\right)$ $\forall j^m \in \left(n_{h-1}^m, n_h^m\right]$.*

[25] See Seppi (1997), Harris (1998), Cordella and Foucault (1999), Goldstein and Kavajecz (2000), and Kadan (2005) for arguments in favor and against the reduction in the tick size in various market structures.

[26] If $\rho \leq \rho^c$, then the spread improvements are so small that the competitive spread is never reached. We discuss this case later.

[27] A closed-form expression for q_0 is given in the proof of this proposition.

Proposition 7 shows that when $\rho > \rho^c$, the equilibria with and without a minimum price variation are qualitatively similar. When $\Delta = 0$, the smallest possible spread is patient traders' *per period* waiting cost, i.e., $\frac{\delta_P}{\lambda}$. In contrast, when $\Delta > 0$, it is equal to this cost *rounded up to* the nearest tick. Thus, not surprisingly, the competitive spread is larger when a minimum price variation is enforced. This *rounding effect* propagates to all equilibrium spreads. To make this statement formal, let $n_h^m(\Delta)$ denote the h^{th} smallest spread in the set of spreads on the equilibrium path when the tick size is $\Delta \geq 0$, and let q_Δ be the number of spreads in this set. The following holds.

Corollary 5. *"Rounding effect": Suppose $\rho > \rho^c$. In equilibrium: (1) $q_\Delta \leq q_0$, (2)$n_h^m(0) \leq n_h^m(\Delta)$, for $h < q_\Delta$, and (3)$n_h^m(0) \leq n_{q_\Delta}^m(\Delta)$ for $q_\Delta \leq h \leq q_0$. This means that the support of possible spreads when the tick size is zero is shifted to the left compared to the support of possible spreads when the tick size is strictly positive.*

Given this result, it is tempting to conclude that the average spread is always minimized when there is no minimum price variation. We show below that this reasoning does not constitute the whole picture because it ignores the impact of the tick size on the dynamics of the spread between transactions.

When $\rho > \rho^c$ and $\delta_P > 0$, in zero-tick equilibrium, traders improve the spread by more than an infinitesimal amount, $\Psi_h^m(0) > 0$.[28] Intuitively, patient traders improve the quote by a non-infinitesimal amount to speed up execution. However, as ρ decreases, spread improvements become smaller and smaller: traders bid less aggressively since market orders arrive more frequently (see the discussion following Proposition 5). When $\Delta > 0$, spread improvements can never be smaller than the tick size; thus, for small values of ρ traders improve prices *by more than they would in the absence of a minimum price variation*. We refer to this effect as being the "*spread improvement effect*." The spread improvement effect works to increase the speed at which spread narrows in between transactions. For this reason, imposing a minimum price variation helps to make the market more resilient. This intuition can be made more rigorous by using the measure of market resiliency, R, defined in Section 2.2.2.

Corollary 6. *Other things being equal, the resiliency of the limit order market, R, is always larger when there is a minimum price variation than in the absence of a minimum price variation. Furthermore, the resiliency of*

[28] Traders must improve upon prevailing quotes (Assumption A.2). However when the tick size is zero, they can improve by an arbitrarily small amount. Proposition 7 shows that they do not take advantage of this possibility when $\rho > \rho^c$.

the market approaches zero as ρ approaches ρ^c from above in the absence of a minimum price variation, whereas it is always strictly greater than zero when a minimum price variation is imposed.

Intuitively, as ρ approaches ρ^c from above, spread improvements become infinitesimal when the spread is large (e.g., equal to K). Thus, the limit orders are submitted arbitrarily close to the largest possible ask price, A, or the smallest possible bid price, B. This explains why, in the absence of a minimum price variation, the resiliency of the market vanishes when ρ goes to ρ^c. Imposing a minimum price variation is a way to avoid this pathological situation, because it forces traders to improve by non-infinitesimal amounts to get price priority.

Thus, intuitively, imposing a minimum price variation can reduce the expected spread, despite the rounding effect, because it makes the market more resilient. We demonstrate this claim by providing a numerical example. The parameter values are as in Example 3 except that $K^m = 0.4375$ ($K = 7$) so that the condition $\rho > \rho^c$ is satisfied ($\rho^c = 0.79$). Table 4 gives all the monetary spreads on the equilibrium path for two different values of the tick size: (1) $\Delta = 0$ and (2) $\Delta = \frac{1}{16}$. The two last lines of the table give the expected spread and the resiliency obtained for each regime. First, observe the "rounding effect"—the five smallest spreads are lower when $\Delta = 0$, than in the case of $\Delta = \frac{1}{16}$. Second, observe the "spread improvement effect"—the spread reduction is quicker for every spread level if a minimum price variation is enforced. This explains why market resiliency is *smaller* when there is no minimum price variation. For this reason, the expected spread turns out to be larger in this case ($0.3675 instead of $0.3265).

So far we have compared a situation with and without a mandatory minimum price variation. More generally, the "spread improvement"

Table 4
Rounding and spread improvement effects

h	$n_h^m(\Delta = 0)$	$n_h^m(\Delta = \frac{1}{16})$
1	$0.05	$0.0625
2	$0.1318	$0.1875
3	$0.1988	$0.3125
4	$0.2535	$0.3750
5	$0.2983	$0.4375
6	$0.3350	*NA*
7	$0.3650	*NA*
8	$0.3896	*NA*
9	$0.4096	*NA*
10	$0.4260	*NA*
11	$0.4375	*NA*
Expected spread	$0.3676	$0.3265
resiliency	3.4×10^{-4}	0.041

Table 5
The tick size minimizing the expected spread

ρ	0.85	0.95	1	1.05	1.15	1.25
Δ^*	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{20}$	$\frac{1}{100}$	$\frac{1}{100}$	$\frac{1}{100}$

effect implies that the expected spread does not necessarily decrease when the tick size is reduced. In order to see this point, consider Table 5. It demonstrates which of the following tick sizes, $\{\frac{1}{100}, \frac{1}{20}, \frac{1}{16}\}$, minimizes the expected spread for different values of ρ. The values of the other parameters are chosen as in Example 3. Consistent with the above argument, the smallest possible tick size $\left(\Delta = \frac{1}{100}\right)$ does not minimize the expected spread for low values of ρ. However as ρ increases, inducing traders to make large improvements by imposing a large minimum price variation becomes less effective, since a high proportion of patient traders induces aggressive limit orders. The "spread improvement effect" becomes of second order compared to the "rounding effect." For this reason, the tick size which minimizes the expected spread becomes smaller.

Finally, we briefly discuss the case $\rho \leq \rho^c$. In this case, traders improve upon large spreads by an infinitesimal amount, posting limit orders arbitrarily close to the largest possible ask price, A, or the smallest possible bid price, B.[29] Thus market resiliency is zero. Imposing a minimum price variation is a way to restore market resiliency since spread improvements are non-infinitesimal as soon as $\Delta > 0$ (Proposition 5).

Notice that the effects associated with a change in the tick size are very similar to those associated with a change in the order arrival rate, λ (Section 3.1). Two forces contribute to a small average spread: (i) *small frictional costs* on the one hand (a small tick and small waiting time between arrivals) and (ii) *large spread improvements*. Our analysis points out that factors which lessen frictional costs may reduce spread improvements, resulting in less resilient markets and eventually higher spreads.

To sum up, reducing or even eliminating the tick size may or may not reduce the average spread. The impact depends on the proportion of patient traders in the market. Many empirical papers have found a decline in the average quoted spreads following a reduction in tick size. These papers, however, do not control for the ratio of patient to impatient traders. In Section 3, we have argued that the proportion of patient traders is likely to decrease over the trading day. In this case, the impact

[29] This would also be the case if patient traders' waiting cost were equal to zero ($\delta_P = 0$). When $\rho \leq \rho^c$ or $\delta_P = 0$, the equilibrium (when there is no minimum price variation) is difficult to describe formally since traders improve upon prevailing quotes by an infinitesimal, but strictly positive, amount.

of a decrease in the tick size on the quoted spread may not be uniform throughout the trading day. Specifically, a decrease in the tick size may increase the average spread at the end of the trading day. To the best of our knowledge, there are no papers testing this hypothesis.

5. Robustness

Recall our assumptions regarding the trading process: A.1—no order cancelations and resubmissions; A.2—limit orders cannot queue at or behind the best quotes; and A.3—buyers and sellers alternate. In this section, we present conditions under which our non-queueing restriction, A.2, is not binding. We also show, using examples, that the main properties of the model persist when we relax the assumption that buyers and sellers alternate. Overall, these robustness tests show that our main results are not driven by the technical assumptions, and that the economic intuitions are still valid when these assumptions are relaxed. For brevity, we omit the proofs of the results given in this section. They can be obtained from the authors upon request.

5.1 Cancelations and resubmissions

Although in practice traders frequently cancel and resubmit their limit orders, our model does not allow them to do so. Hence, it cannot explain why traders actively manage their orders. Clearly, modeling cancelations and resubmissions is important. But it is also very difficult. One possible approach, followed by Hollifield *et al.* (2003) or Goettler, Parlour, and Rajan (2003) assumes that cancelations occur exogenously at random points in time. In our model, all orders must eventually be executed. Thus, in order to follow this approach, we would need to *arbitrarily* specify the payoff to a trader when his order is canceled. For this reason, we do not engage in this exercise.[30]

Most cancelations appear as the result of a particular behavior that we do not seek to capture in this study. Hasbrouck and Saar (2002) find that the majority of cancelations on Island ECN occurs very quickly after order submission (about 60% are canceled within 30 seconds). They argue that these "fleeting" orders seek liquidity rather than provide it. Tkach (2002) studies the limit order submission in the 100 most liquid stocks on the Tel Aviv Stock Exchange. She shows that the median time to cancelation is 11 minutes, and over 12% of all cancelations occur within a minute of submission. This is a short period of time, especially given the low volume in many of these stocks. We do not expect that these cancelation

[30] Rosu (2004) considers a dynamic model of price formation in limit order markets. As in our model, traders value speed of execution. Assuming that limit orders resubmission is costless and instantaneous, he allows traders to cancel and resubmit their orders. Some of his results are qualitatively similar to ours. In particular, he finds that resiliency increases with the proportion of patient traders.

strategies affect our conclusions. They are phenomena outside the scope of this model.

5.2 Queuing at the inside quotes

We assume that traders cannot place limit orders at or behind the inside quotes. In fact, such orders are allowed and used.[31] Queuing, however, should not invalidate the findings that (i) an increase in the proportion of patient traders or (ii) a decrease in the order arrival rate yield more resilient and more competitive limit order markets. Actually, limit order traders' incentive to jump ahead of the queue is greater when time to execution increases, that is, when the proportion of patient traders rises or the order arrival rate declines. This means that spreads narrow more rapidly in these two cases, even when queuing is an option.

This reasoning suggests that if the proportion of patient traders is sufficiently large or the order arrival rate is sufficiently small then traders will choose not to queue. In this case, the equilibrium is exactly as described in Section 2. This result is established in the next proposition.

Proposition 8. *Suppose traders are heterogeneous and are allowed to queue at the inside quotes subject to time priority (i.e., limit orders entered first at a given price are executed first). If*

$$\frac{\lambda\Delta}{\delta_P} \leq 2[1 + \theta_P(2 - \theta_P)], \tag{15}$$

then the equilibrium when traders are not allowed to queue is an equilibrium in this setting.

Suppose that traders use the order placement strategies described in Section 2 and give them the freedom to queue at the best quotes. Under condition (15), traders prefer to submit limit orders improving upon the inside quotes rather than queuing. Hence, traders' strategies form an equilibrium even though traders have the possibility to queue. Condition (15) is satisfied in all the numerical examples in the article. It follows that the possibility of queuing does not *per se* invalidate our comparative statics.

Inequality (15) is satisfied when the order arrival rate is sufficiently small. Furthermore, a rise in the proportion of patient traders increases the right-hand side of Inequality (15) and thereby helps to satisfy this

[31] For instance, Biais, Hillion, and Spatt (1995) report that about 5% of new buy limit orders are placed at the best bid price. They find a similar frequency for sell limit orders.

inequality. This means that traders are less likely to queue when the order arrival rate is small and the proportion of patient traders is large, as conjectured. Finally, Inequality (15) is satisfied when the tick size, Δ, is sufficiently small. In this case, queuing is never optimal because liquidity providers can jump ahead of the queue at a low cost. This reasoning suggests that the number of limit orders placed at the same price should have decreased following tick size reductions.

5.3 Buyers and sellers arrive randomly

We have assumed that buyers and sellers alternate. Clearly, it would be more realistic to assume that the sequence of arrivals for buyers and sellers is random. Unfortunately, in this case, the model becomes intractable for two reasons. First, it is not possible to solve for the equilibrium by induction because the waiting time function is not recursive. Second, the waiting time for a limit order depends on the state of the limit order book when the order is placed and not simply on the inside spread.

Under these conditions, it is very difficult to compute the equilibrium analytically. Such a computation involves the following steps. First, conjecture equilibrium order placement strategies for patient and impatient traders. Second, use the conjectured equilibrium strategies to calculate the expected waiting time for each possible limit order in each possible state of the book. This task requires solving a number of simultaneous linear equations which grows quickly with K because the waiting time function depends on the entire state of the book. Third, check that the "conjectured" strategies are optimal given the expected waiting times computed in the second step. If these strategies are not optimal, the steps are repeated until an equilibrium is found.

This procedure is tedious even for small values of K. It can be implemented for specific values of the parameters, however. Thus, we use examples to demonstrate that the economic intuitions of our model persist when buyers and sellers arrive randomly. We assume that each trader is a buyer or a seller with equal probabilities. We focus on the case $K = 4$. This choice allows for different levels of spread improvements. For example, when the spread has four ticks, a limit order trader can improve upon prevailing quotes by one tick (small improvement), two or three ticks (large improvements). We consider 3 different sets of values for the parameters. In Example 4, the parameters are such that traders are homogeneous while in Examples 5 and 6, the parameters are such that traders are heterogeneous. The proportion of patient traders is larger in Example 5 than in Example 6. We describe below the equilibrium obtained in each example.[32]

[32] A detailed derivation of the claims in these examples can be obtained from the authors upon request.

5.3.1 Example 4—A Strongly Resilient Book (homogenous traders). Set $K = 4$, $\Delta = \frac{1}{16}$, $\delta_P = \delta_I = 0.025$ (traders are homogeneous). The following order placement strategy constitutes an equilibrium: (i) when the spread is larger than one tick, buyers and sellers of both types submit a 1-limit order and (ii) when the spread is equal to one tick, both submit a market order. Following a transaction, the spread increases to four ticks, but then reverts to the traders' reservation spread of one tick before the next transaction. This market is therefore strongly resilient: $R = 1$. As traders are homogeneous, the equilibrium is not affected by θ_P, the proportion of patient traders.

5.3.2 Example 5—A Resilient Book (heterogenous traders, large $\boldsymbol{\theta_P}$). Set: $\Delta = \frac{1}{16}$, $K = 4$, $\theta_P = 0.7$, $\lambda = 1$, $\delta_P = 0.01$, and $\delta_I = 0.07$. In this case traders are heterogeneous. The following order placement strategies constitute an equilibrium. An impatient trader always submits a market order. A patient trader submits (i) a 2-limit order when the spread is equal to three or four ticks, (ii) a 1-limit order when the spread is equal to two ticks, and (iii) a market order when the spread is equal to one tick. The resiliency of the market is $R = 0.49$.

5.3.3 Example 6—A Weakly Resilient Book (heterogenous traders, small $\boldsymbol{\theta_P}$). Set: $\Delta = \frac{1}{16}$, $K = 4$, $\theta_P = 0.3$, $\lambda = 1$, $\delta_P = 0.01$, and $\delta_I = 0.07$. The following order placement strategies constitute an equilibrium. An impatient trader always submits a market order. When the spread is larger than one tick, a patient trader places a limit order improving the spread by one tick. When the spread is equal to one tick, a patient trader places a market order. The resiliency of the market is $R = 0.027$.

Clearly, the equilibrium obtained in each example has the same properties as the equilibrium obtained when buyers and sellers alternate. In Example 4, the spread oscillates between a large level and a small level. This is expected since this pattern epitomizes the homogeneous case (see Proposition 1). Furthermore, as in the baseline model, limit order traders use a more aggressive bidding strategy when the proportion of patient traders is large (i.e., in Example 5). To see this, consider the case in which the spread is equal to four ticks and a patient trader arrives in the market. In Example 5, the trader improves upon prevailing quotes by two ticks whereas in Example 6 he improves by only one tick. The economic intuition is exactly the same as when buyers and sellers alternate. Limit order traders bid more aggressively when θ_P is large because their waiting times are larger, other things being equal. It follows that the resiliency of the market increases in the proportion of patient traders. Accordingly, we find that the stationary distribution of spreads is skewed toward small spreads in Example 5 and large spreads in Example 6. For instance, the probability that the inside spread is equal to four ticks is 4% in Example 5 and 38% in Example 6. Again, this is expected since the proportion of patient traders is larger in Example 5.

These findings suggest that relaxing the alternating arrival assumption does not change the conclusions obtained when buyers and sellers alternate. The driving force of our model is that limit order traders react to exogenous increases in their total waiting costs by submitting more aggressive orders. This basic economic intuition does not hinge on the assumption that buyers and sellers alternate. However, relaxing this assumption prevents us from solving the model in general. We view our model as a way to bypass this problem without losing much of the economic intuitions.

6. Conclusion

We construct a model of price formation in a limit order market. Agents in our model are strategic, trade for liquidity reasons, and differ in their impatience. Upon arrival, they must decide on whether to submit a market order or a limit order. Their choice is driven by a trade-off between the cost of immediacy (the spread) and the cost of delayed execution, as first suggested by Demsetz (1968).

We derive the equilibrium order placement strategies. We find that the proportion of patient traders in the population and the order arrival rate are the key determinants of the limit order book dynamics. Traders submit aggressive limit orders (improve upon quoted spreads by large amounts) when the proportion of patient traders is large or when the order arrival rate is low. For this reason, markets with a high proportion of patient traders or a small order arrival rate are more resilient. Also, a reduction in the tick size reduces market resiliency, and in some cases increases the average spread.

The analysis yields several testable predictions: (i) a positive relationship between inter-trade durations (conditional on the spread) and market resiliency; (ii) a negative relationship between the order arrival rate and market resiliency; (iii) a joint decline of limit order aggressiveness and market resiliency at the end of trading sessions; and (iv) limit order traders submit more (respectively less) aggressive orders when the spread is large if patient (respectively impatient) traders dominate the trading population.

Future research will focus on relaxing some assumptions that limit the scope of our results. We assume that the proportion of patient traders in a given market is exogenous. It would be interesting to endogenize the composition of the trading population to gain insights on the sources of cross-sectional variations in this composition. The order arrival rate is independent of the state of the limit order book in our model. In practice, traders time their arrivals during the day according to market conditions, and variations in the size of the inside spread may then trigger changes in the intraday order arrival rate. This relationship is not captured by our model. Finally, we have observed that the equilibrium outcome is in general inefficient in our model. This result raises the possibility that introducing designated intermediaries in order driven markets could be efficiency enhancing, pointing to another interesting direction for future work.

Appendix

Proof of Lemma 1.
Step 1. Suppose a trader (say a buyer) submits a *j*-limit order when the spread is *s*. By A.3 the following trader is a seller. We claim that at the time the *j*-limit order is cleared, the spread will revert to *s*. We prove this claim by induction on *j*. If *j* = 1 then by A.2 the next order is a sell market order, and the spread immediately reverts to *s*. Suppose now that *j* > 1, and assume that our assertion is true for all $k = 1, \ldots, j-1$. By A.2, the seller must either submit a market order or submit a *k*-limit order with $k = 1, \ldots, j-1$. If the seller submits a market order, then the spread reverts *s*. If, on the other hand, the seller submits a *k*-limit order with $k \in \{1, \ldots, j-1\}$, then by the induction hypothesis, when that seller's *k*-limit order is cleared the spread reverts to *j*. It follows that when the *j*-limit order is cleared, the spread reverts to *s* as required.
Step 2. Consider a trader, say a buyer, who submits a *j*-limit order. The expected waiting time of this order from this moment on is *T(j)*. By A.2, this buyer acquires price priority (he posts the best bid price). Suppose that the next trader (a seller by A.3) submits a *k*-limit order with $k \in \{1, \ldots, j-1\}$. When this *k*-limit order will be executed, the spread will revert to *j* (Step 1). As traders do not cancel their orders or do not submit orders behind the best quotes, the state of the book will then be exactly as when the buyer initially posted the *j*-limit order. In particular, the buyer will have price priority. Thus, when the spread reverts to *j*, the buyer's expected waiting time from that moment on is *T(j)* as well.
Step 3. We have explained in the text why $T(1) = \frac{1}{\lambda}$. Now, consider a trader (say a buyer) who submits a *j*-limit order with $j > 1$. The next trader (a seller) must choose among *j* options. With probability $\alpha_0(j)$, he submits a market order that clears the buyer's limit order. In this case, the expected waiting time of the buyer is $\frac{1}{\lambda}$. With probability $\alpha_k(j)$, the seller submits a *k*-limit order ($k = 1, \ldots, j-1$). In the latter case, the original buyer's expected waiting time is $\frac{1}{\lambda} + T(k) + T(j)$. Indeed, he has to wait (1) $1/\lambda$—for the seller to arrive, (2) $T(k)$—until the seller's order is cleared and the spread reverts to *j* (by Step 1), and (3) another *T(j)* as we are back to the original position (by Step 2). Overall the original buyer's expected waiting time, *T(j)*, is given by:

$$T(j) = \frac{\alpha_0(j)}{\lambda} + \sum_{k=1}^{j-1} \alpha_k(j) \left[\frac{1}{\lambda} + T(k) + T(j) \right]. \tag{16}$$

If $\alpha_0(j) > 0$, we obtain the second part of the lemma by solving for *T(j)* and using the fact that $\sum_{k=0}^{j-1} \alpha_k(j) = 1$. As for the third part of the lemma: If $\alpha_0(j) = 0$, then the seller never submits a market order when the spread is *j*. Thus, the waiting time of the buyer who creates the *j*-limit order is infinite: $T(j) = +\infty$. ■

Proof of Proposition 1. It follows immediately from the arguments preceding the proposition. ■

Proof of Proposition 2. We first prove the following lemma.

Lemma 3. Suppose that facing a spread of size s ($s \in \{1, \ldots, K-1\}$), trader i ($i \in \{P, I\}$) submits a j-limit order with $0 \leq j < s$. Then, facing a spread of size s +1, he either submits an s-limit order or submits a j-limit order.

Proof. By assumption, trader *i* submits a *j*-limit order when he faces a spread of size *s*. Thus:

$$\pi_i(j) \geq \pi_i(k) \quad k = 0, \ldots j-1, j+1, \ldots, s-1.$$

Now, suppose that trader *i* faces a spread of size $s + 1$. If $\pi_i(s) < \pi_i(j)$, then trader *i* will submit a *j*-limit order since $\pi_i(j) \geq \pi_i(k)$ for all $k = 0, \ldots, s$. If $\pi_i(s) \geq \pi_i(j)$, then trader *i* submits a *s*-limit order since $\pi_i(s) \geq \pi_i(j) \geq \pi_i(k)$ for all $k = 0, \ldots, s-1$. ■

By definition of the reservation spread, and since $\delta_P < \delta_I$, it follows that:

$$\pi_I(j) < \pi_P(j) < 0, \forall j < j_P^*.$$

Thus, all traders submit a market order when they face a spread which is smaller than or equal to patient traders' reservation spread. This implies that $T^*(1) = T^*(2) = \ldots T^*\left(j_P^*\right) = \frac{1}{\lambda}$. Now suppose a patient trader faces a spread of size $j_P^* + 1$. Lemma 3 implies that he will either submit a j_P^*-limit order or submit a market order. He obtains a larger payoff with a j_P^*-limit order since

$$\pi_P\left(j_P^*\right) = j_P^*\Delta - T^*\left(j_P^*\right)\delta_P = j_P^*\Delta - \frac{\delta_P}{\lambda} \geq 0,$$

where the last inequality follows from the definition of j_P^*. Then, we deduce from Lemma 3 that the patient type submits limit orders for all spreads $s \in \left\langle j_P^* + 1, K\right\rangle$. As for the impatient type there are two cases:

Case 1: The impatient type submits a market order for each $s \in \left\langle j_P^* + 1, K\right\rangle$ in which case we set $s_c = K$.

Case 2: There are spreads in $\langle 1, K\rangle$ for which the impatient type submits limit orders. In this case, let s_c be the smallest spread that an impatient trader creates with a limit order. By definition of s_c, the impatient trader submits a market order when he faces a spread $s \in \langle 1, s_c\rangle$ and a s_c-limit order when he faces a spread of size $s_c + 1$. Then, we deduce from Lemma 3 that impatient traders submit a limit order when they face a spread in $\langle s_c + 1, K\rangle$ and a market order otherwise. Finally, it cannot be optimal for an impatient trader to submit a limit order which creates a spread smaller than his reservation spread. This implies $s_c \geq j_I^*$. ■

Proof of Proposition 3. Since we assume that $s_c = K$, the impatient type always submits market orders. From Proposition 2, a patient trader submits a market order when he faces a spread in $\left\langle 1, j_P^*\right\rangle$ and a j_P^*-limit order when he faces a spread of size $j_P^* + 1$. Repeated application of Lemma 3 (see the proof of Proposition 2) shows the existence of spreads $n_1 < n_2 < \ldots < n_q$ such that facing a spread in $\langle n_h + 1, n_h + 1\rangle$ the patient trader submits an n_h-limit order for $h = 1, \ldots, q-1$. Clearly, $n_1 = j_P^*$ and $n_q = K$. ■

Proof of Proposition 4. When they observe a spread of size n_1, all the traders submit a market order. Therefore $T^*(n_1) = \frac{1}{\lambda}$. Let $h \in \{2, \ldots, q\}$. Suppose that the posted spread is $s \in \langle n_{h-1} + 1, n_h\rangle$. When he observes this spread, a patient trader submits an n_{h-1}-limit order, and an impatient trader submits a market order (Proposition 3). Therefore when the posted spread is $s \in \langle n_{h-1}+1, n_h \rangle$, we have $\alpha_0(s) = 1-\theta_P, \alpha_{n_{h-1}}(s) = \theta_P$, and $\alpha_k(s) = 0, \forall\, k \notin \{0, n_{h-1}\}$. Thus, Lemma 1 (second part) yields

$$T^*(s) = \frac{1}{1-\theta_P}\left[\frac{1}{\lambda} + \theta_P T^*(n_{h-1})\right],\ \forall s \in \langle n_{h-1}+1,\, n_h \rangle. \tag{17}$$

Hence, $T^*(\cdot)$ is constant for all $s \in \langle n_{h-1}+1, n_h \rangle$. Using Equation (17), we obtain

$$T^*(n_{h+1}) - T^*(n_h) - \rho(T^*(n_h) - T^*(n_{h-1})) \quad for \quad h \geq 2, \tag{18}$$

where $\rho = \frac{\theta_P}{\theta_I}$. Furthermore, using Equation (17) and the fact that $T^*(n_1) = \frac{1}{\lambda}$, we obtain

$$T^*(n_2) - T^*(n_1) = \frac{2\rho}{\lambda} > 0.$$

The claim follows now by repetitive application of Equation (18) and from the fact that $T^*(n_1) = \frac{1}{\lambda}$. ■

Proof of Proposition 5. Since $n_h = n_{h-1} + \Psi_h$, we immediately get that $n_h = n_1 + \sum_{k=2}^{h} \Psi_k$. Furthermore, since $n_q = K$, it must be the case that q is the smallest integer such that $n_1 + \sum_{k=2}^{q} \Psi_k \geq K$. The expression for Ψ_h is given by Equation (8). ■

Proof of Corollary 1. Recall that q is the smallest integer such that $n_1 + \sum_{k=2}^{q} \Psi_k \geq K$. It follows that q (a) decreases with θ_P and δ_P and (b) increases with λ, since Ψ_k increases with θ_P and δ_P and decreases with λ, for all $k \in \{1, \ldots, q-1\}$. The result is then immediate. ■

Proof of Lemma 2. We first show that the Markov chain given by W is (a) irreducible and (b) a-periodic.

The Markov chain is irreducible. Observe that given any two states j_1, j_2 with $1 \leq j_1 < j_2 \leq q$ there is a positive probability that the chain will move from j_1 to j_2 after a sufficiently large (though finite) number of transitions. This implies that any two states in the chain communicate; hence, the chain is irreducible.

The Markov chain is a-periodic. Notice that $W_{q,q} = \theta_I > 0$. This means that when the chain is in state q, there is a probability equal to θ_I^n that it will stay in this state for the next n transitions, $\forall n \geq 1$. Since state q communicates with all the other states of the chain, it follows that no state has a period greater than 1. Thus the chain is a-periodic.

These properties imply that the Markov chain is ergodic. Hence, the induced Markov chain has a unique stationary probability distribution of spreads (see Feller 1968). Let $u = (u_1, \ldots, u_q)$ denote the row vector of stationary probabilities. The stationary probability distribution is obtained by solving q + 1 linear equations given by:

$$uW = u \text{ and } u\varepsilon = 1, \tag{19}$$

where ε stands for the unit column vector. It is straightforward to verify that the probabilities given by Equation (10) are a solution of this system of equations. ■

Proof of Proposition 6. The proof follows immediately from Equation (11). ■

Proof of Corollary 2. In the proof of Corollary 1, we have established that q increases with λ. Thus $q_{\lambda_S} \leq q_{\lambda_F}$. Using Proposition 5, we obtain

$$n_{k+1}(\lambda) = n_k(\lambda) + CF\left(\frac{2\rho^{h-1}\delta_P}{\lambda\Delta}\right), \text{ for } \lambda \in \{\lambda_S,\lambda_F\} \text{ and } k \leq q_{\lambda_S} - 2.$$

Thus, if $n_k(\lambda_F) \leq n_k(\lambda_S)$, then $n_{k+1}(\lambda_F) \leq n_{k+1}(\lambda_S)$ for $k \leq q_{\lambda_S} - 2$. Now, observe that for $k = 1$, we have (using Proposition 5):

$$n_1(\lambda) = CF\left(\frac{\delta_P}{\lambda\Delta}\right).$$

We deduce that $n_1(\lambda_F) \leq n_1(\lambda_S)$ and conclude that $n_k(\lambda_F) \leq n_k(\lambda_S)$ for $k \leq q_{\lambda_S} - 1$. Furthermore, $n_{q_{\lambda_S}}(\lambda_S) = n_{q_{\lambda_F}}(\lambda_F) = K$. Consequently, $n_k(\lambda_F) \leq n_{q_{\lambda_S}}(\lambda_S)$ for $q_{\lambda_S} \leq k \leq q_{\lambda_F}$. This proves the first part of the corollary. The second part follows from Corollary 1. ■

Proof of Corollary 3. Let $\widetilde{N}_h$ denote the random variable describing the number of trader arrivals between two consecutive transactions, conditional on the event that the first transaction took place when the spread was n_h. The conditional duration is:

$$\bar{D}_h = \frac{E\left(\widetilde{N}_h\right)}{\lambda} \quad \forall h = 1, \ldots, q, \tag{20}$$

since the expected waiting time between two order arrivals is $\frac{1}{\lambda}$. Now we compute $E\left(\widetilde{N}_h\right)$. Suppose that the last transaction took place at the smallest possible spread, n_1. Following this transaction, the new spread in equilibrium is n_2. If the next trader is an impatient trader, then a new transaction takes place and $N_1 = 1$. If the next trader is a patient trader, he submits a limit order which creates a spread equal to n_1. Then, the next order is a market order since all traders submit market orders when the spread is n_1. In this case $N_1 = 2$. We deduce that the probability distribution for $\widetilde{N}_1$ is:

$$\Pr(N_1 = 1) = \theta_I \text{ and } \Pr(N_1 = 2) = \theta_P.$$

More generally, the same type of reasoning yields the probability distribution for $\widetilde{N}_h$ when $1 \leq h < q$. The largest possible value for $\widetilde{N}_h$ is $h + 1$ and

$$\Pr(N_h = j) = \theta_I\theta_P^{j-1} \quad \text{for} \quad j = 1, \ldots, h,$$

and

$$\Pr(N_h = h + 1) = \theta_P^h.$$

We deduce that

$$E\left(\widetilde{N}_h\right) = \theta_I \sum_{j=1}^{h} j\theta_P^{j-1} + (h + 1)\theta_P^h,$$

which simplifies as

$$E\left(\tilde{N}_h\right) = \frac{1-\theta_P^{h+1}}{\theta_I}, \text{ for } 1 \leq h < q. \tag{21}$$

The expression for $\bar{D}_h$ follows from this equation and Equation (20). Finally, observe that when the last transaction takes place at the largest possible spread, n_q then the spread following this transaction remains n_q. Hence, the situation is as if the last transaction took place at spread, n_{q-1}. It follows that the probability distributions of $\tilde{N}_q$ and $\tilde{N}_{q-1}$ are identical. Therefore $E\left(\tilde{N}_q\right) = E\left(\tilde{N}_{q-1}\right)$. The expression for $\bar{D}_q$ follows. The last part of the proposition follows directly from the expression for $\bar{D}_h, h \leq q$. ■

Proof of Corollary 4. The size of spread improvement (in number of ticks) when the current spread is n_h is given by $\Psi_h = CF\left(\frac{2\rho^{h-1}\delta_P}{\lambda\Delta}\right)$. Thus when $\rho < 1, \Psi_h$ decreases with h and when $\rho > 1, \Psi_h$ increases with h. This means that when $\rho < 1$, spread improvements are inversely related to the inside spreads on the equilibrium path. In contrast, when $\rho > 1$, spread improvements are positively related to the inside spreads on the equilibrium path. ■

Proof of Proposition 7.

Step 1. We first derive the expected waiting time function associated with the order placement strategies described in Parts 1, 2, and 3 of the proposition. All traders submit a market order when they face a spread equal to n_1^m. It follows that $T^*\left(n_1^m\right) = \frac{1}{\lambda}$. Now suppose that the posted spread is $s^m \in \left(n_{h-1}^m, n_h^m\right]$ with $h \geq 2$. When he observes this spread, a patient trader submits an n_{h-1}^m-limit order and an impatient trader submits a market order. Therefore $\alpha_0(s^m) = \theta_I$ and $\alpha_{n_{h-1}}(s^m) = \theta_P$. It follows that

$$T^*(s^m) = \frac{\theta_I}{\lambda} + \theta_P\left(\frac{1}{\lambda} + T^*\left(n_{h-1}^m\right) + T^*(s^m)\right), \forall s^m \in \left(n_{h-1}^m, n_h^m\right] \text{ for } h \geq 2, \tag{22}$$

which yields

$$T^*(s^m) = \frac{1}{\theta_I}\left[\frac{1}{\lambda} + \theta_P T^*\left(n_{h-1}^m\right)\right], \forall s^m \in \left(n_{h-1}^m, n_h^m\right] \text{ for } h \geq 2.$$

Hence $T^*(\cdot)$ is constant for all $s^m \in \left(n_{h-1}^m, n_h^m\right]$ with $h \geq 2$. Then, following the last part of the proof of Proposition 4, it is straightforward to show that the expected waiting time function is:

$$T^*\left(n_h^m\right) = \frac{1}{\lambda}\left[1 + 2\sum_{k=1}^{h-1}\rho^k\right] \forall \quad h = 2,\ldots, q_0 - 1.$$

This proves the last part of the proposition. Observe that the last equation implies:

$$\left(T^*\left(n_h^m\right) - T^*\left(n_{h-1}^m\right)\right)\delta_P = 2\rho^{h-1}\frac{\delta_P}{\lambda} = n_h^m - n_{h-1}^m, \tag{23}$$

where the second equality follows from Part 3 of the proposition.

Step 2. Now we show that the order placement strategies described in Parts 1, 2, and 3 of the proposition are optimal given the expression of the waiting time function given in Part 4. Equation (23) implies that:

$$n_h^m - T^*\left(n_h^m\right)\delta_P = n_{h-1}^m - T^*\left(n_{h-1}^m\right)\delta_P = \ldots = n_1^m - T^*\left(n_1^m\right)\delta_P \text{ for } \quad h = 2,\ldots, q_0 - 1. \tag{24}$$

As $n_1^m = \frac{\delta_P}{\lambda}$ and $T^*\left(n_1^m\right) = \frac{1}{\lambda}$, we have

$$n_1^m - T^*\left(n_1^m\right)\delta_P = 0. \tag{25}$$

Now Equations (24) and (25) imply that:

$$n_h^m - T^*\left(n_h^m\right)\delta_P = 0 \text{ for } \quad h = 1,...q_0 - 1. \tag{26}$$

Furthermore, we know (see Step 1) that $T^*(\cdot)$ is constant for all $s^m \in \left(n_{h-1}^m, n_h^m\right]$ with $h \geq 2$:

$$T^*(s^m) = T^*\left(n_h^m\right) \text{ for } \quad s^m \in \left(n_{h-1}^m,\, n_h^m\right] \quad \text{and} \quad h = 2,...,\, q_0,$$

which implies that for $h \in \{2, ..., q_0\}$:

$$s^m - T^*(s^m)\delta_i < n_h^m - T^*\left(n_h^m\right)\delta_i \text{ for } \quad s^m \in \left(n_{h-1}^m,\, n_h^m\right) \quad \text{and} \quad i \in \{P,I\}. \tag{27}$$

Consider an impatient trader who faces a spread with size $j^m \in \left(n_{h-1}^m, n_h^m\right]$. Note that the spread can be on the equilibrium path $\left(j^m = n_h^m\right)$ or not $\left(j^m < n_h^m\right)$. As $\delta_P < \delta_I$, Equation (26) implies

$$n_k^m - T^*\left(n_k^m\right)\delta_I < 0 \text{ for } \quad k = 1,...h - 1.$$

Using Equation (27), we deduce that

$$s^m - T^*(s^m)\delta_I < 0, \forall s^m \in (0, j^m).$$

Thus any limit order yields a negative payoff to the impatient trader. It follows that he submits a market order (which has a zero payoff).

Now consider a patient trader who faces a spread with size $j^m \in \left(n_{h-1}^m, n_h^m\right]$, with $h > 2$. From Equations (26) and (27), we know that:

$$s^m - T^*(s^m)\delta_P < 0, \forall s^m \notin \left\{0,\, n_1^m,\, n_2^m,...,\, n_{h-1}^m\right\}. \tag{28}$$

Furthermore we know from Equation (26) that:

$$n_k^m - T^*\left(n_k^m\right)\delta_P = 0, \quad for \quad k \in \{1,2,...,\, h-1\}. \tag{29}$$

We deduce from Equations (28) and (29) (i) that the patient trader's best response belongs to the set $\left\{n_1^m, n_2^m, ..., n_{h-1}^m\right\}$ and (ii) that he is indifferent between any spread in this set. Thus, it is a best response to choose n_{h-1}^m for the patient trader when he faces a spread with size $j^m \in \left(n_{h-1}^m, n_h^m\right]$. Now consider a patient trader who faces a spread with size $j^m \leq n_1^m$. It follows from Equation (28) that the patient trader cannot profitably improve upon this spread. Therefore he chooses a market order.

Step 3. Finally, we compute the expression for q_0. Since $n_h^m = n_{h-1}^m + \Psi_h^m(0)$, we immediately get that $n_h^m = n_1^m + \sum_{k=2}^h \Psi_k^m(0)$. Furthermore since $n_{q_0} = K^m$, it must be the case that q_0 is the smallest integer such that $n_1^m + \sum_{k=2}^{q_0} \Psi_k^m(0) \geq K^m$. As $\Psi_k^m(0) = \frac{(2\rho^{k-1})\delta_P}{\lambda}$, we deduce that q_0 is the smallest integer such that:

$$\frac{\delta_P}{\lambda} + \sum_{k=2}^{q_0} \frac{(2\rho^{k-1})\delta_P}{\lambda} \geq K^m. \tag{30}$$

Now the smallest integer q_0 which satisfies Condition (30) is given by:

$$q_0 = \begin{cases} CF\left(\frac{\ln[(\rho-\rho^c)(\frac{K^m\lambda+\delta_P}{2\delta_P})]}{\ln(\rho)}\right) & \text{if}\,\rho \neq 1 \quad \text{and} \quad \rho > \rho^c, \\ CF\left(\frac{K^m\lambda+\delta_P}{2\delta_P}\right) & \text{if} \quad \rho = 1. \end{cases} \tag{31}$$

There is no finite solution if $\rho \leq \rho^c$. Using the definition of ρ^c and the fact that $K^m > \frac{\delta_P}{\lambda}$, it is straightforward to check that $q_0 \geq 2$. This achieves the proof of Proposition 7. ■

Proof of Corollary 5. Using Propositions 5 and 7, we obtain

$$n^m_{k+1}(0) = n^m_k(0) + \frac{2\rho^{h-1}\delta_P}{\lambda},$$

and

$$n^m_{k+1}(\Delta) = n^m_k(\Delta) + CF\left(\frac{2\rho^{h-1}\delta_P}{\lambda\Delta}\right)\Delta,$$

for $1 \leq k \leq \min\{q_0 - 2, q_\Delta - 2\}$. Thus if $n^m_k(0) < n^m_k(\Delta)$, then $n^m_{k+1}(0) < n^m_{k+1}(\Delta)$ for $1 \leq k \leq Min\{q_0 - 2, q_\Delta - 2\}$. Now observe that for $k = 1$, we have (using Propositions 5 and 7):

$$n^m_1(0) = \frac{\delta_P}{\lambda} \quad \text{and} \quad n^m_1(\Delta) = CF\left(\frac{\delta_P}{\lambda\Delta}\right)\Delta.$$

Hence $n^m_1(0) < n^m_1(\Delta)$ since $\frac{\delta_P}{\lambda\Delta} < CF(\frac{\delta_P}{\lambda\Delta})$. We deduce that $n^m_k(0) < n^m_k(\Delta)$ for $k \leq Min\{q_0 - 1, q_\Delta - 1\}$. Recall that q_0 and q_Δ are the smallest integers such that:

$$n^m_{q_0-1}(0) + 2\rho^{q_0-1}\frac{\delta_P}{\lambda} \geq K^m \quad \text{and} \quad n^m_{q\Delta-1}(\Delta) + CF\left(2\rho^{q\Delta-1}\frac{\delta_P}{\lambda\Delta}\right)\Delta \geq K^m$$

Since $n^m_{q_0-1}(0) < n^m_{q_\Delta-1}(\Delta)$, we deduce that $q_\Delta \leq q_0$. Thus we have proved Parts 1 and 2 of the corollary. The last part is straightforward since $n^m_{q_\Delta}(\Delta) = K\Delta = K^m = n^m_{q_0}(0)$. ■

Proof of Corollary 6. Recall that we measure resiliency by $R = \theta_P^{q-1}$. As $\rho = \frac{\theta_P}{\theta_I}$, we can also write this measure in function of ρ: $R = \left(\frac{\rho}{1+\rho}\right)^{q-1}$. Let $R(\Delta, \rho)$ be the value of resiliency for a given tick size, Δ and a given value of the ratio ρ. In Corollary 5, we have shown that $q_\Delta \leq q_0$. We deduce that $R(\Delta, \rho) \geq R(0, \rho)$. Using the expression for q_0 given in Equation (31) (see proof of Proposition 7), it is readily shown that $lim_{\rho\to\rho^c} q_0 = \infty$. It follows that $lim_{\rho\to\rho^c} R(0, \rho) = 0$. When $\Delta > 0$, the number of spreads on the equilibrium path cannot be larger than K, that is $q\Delta \leq K$. We deduce that $R(\Delta, \rho) \geq \left(\frac{\rho}{1+\rho}\right)^{K-1} > 0$ for $\Delta > 0$. ■

References

Admati A., and P. Pfleiderer, 1988, “A Theory of Intraday Patterns: Volume and Price Variability,” *Review of Financial Studies*, 1, 3–40.

Angel, J., 1994, “Limit versus Market Orders,” working paper, Georgetown University.

Beneish, M., and R. Whaley, 1996, “An Anatomy of the ‘S&P Game’: The Effects of Changing Rules,” *Journal of Finance*, 51, 1909–1930.

Biais, B., P. Hillion, and C. Spatt, 1995, “An Empirical Analysis of the Limit Order Book and the Order Flow in the Paris Bourse,” *Journal of Finance*, 50, 1655–1689.

Biais, B., D. Martimort, and J. C. Rochet, 2000, "Competing Mechanism in a Common Value Environment," *Econometrica*, 68, 799–837.

Black, F., 1971, "Towards a Fully Automated Exchange, Part I," *Financial Analysts Journal*, 27, 29–34.

Bloomfield R., M. O'Hara, and G. Saar, 2002, "The 'Make or Take' Decision in Electronic Markets: Evidence on the Evolution of Liquidity," *Journal of Financial Economics*, 55, 425–459.

Chakravarty, S., and C. Holden, 1995, "An Integrated Model of Market and Limit Orders," *Journal of Financial Intermediation*, 4, 213–241.

Chung, K., B. Van Ness, and R. Van Ness, 1999, "Limit Orders and the Bid-Ask Spread," *Journal of Financial Economics*, 53, 255–287.

Coppejans, M., I. Domowitz, and A. Madhavan, 2003, "Dynamics of Liquidity in an Electronic Limit Order Book Market," working paper, Duke University.

Cordella, T., and T. Foucault, 1999, "Minimum Price Variations, Time Priority and Quote Dynamics," *Journal of Financial Intermediation*, 8, 147–173.

Degryse, H., F. deJong, M. Ravenswaaij, and G. Wuyts, 2003, *Aggressive Orders and the Resiliency of a Limit Order Market*, Mimeo, Leuven University.

Demsetz, H., 1968, "The Costs of Transacting," *Quarterly Journal of Economics*, 82, 33–53.

Domowitz, I., and A. Wang, 1994, "Auctions as Algorithms," *Journal of Economic Dynamics and Control*, 18, 29–60.

Easley, D., and M. O'Hara, 1992, "Time and the Process of Security Price Adjustment," *Journal of Finance*, 47, 577–605.

Engle, R., and A. Patton, 2003, "Impacts of Trades in an Error-Correction Model of Quote Prices," *Journal of Financial Markets*, 7, 1–25.

Engle, R., and J. Russel, 1998, "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data," *Econometrica*, 66, 1127–1162.

Foster, F., and S. Viswanathan, 1993, "Variations in trading Volume, Return Volatility and Trading Costs: Evidence on Recent Price Formation Models," *Journal of Finance*, 48, 187–211.

Feller, W., 1968, *An Introduction to Probability Theory and its Applications* (3rd ed.), Wiley, New York.

Foucault, T., 1999, "Order Flow Composition and Trading Costs in a Dynamic Limit Order Market," *Journal of Financial Markets*, 2, 99–134.

Glosten, L., 1994, "Is the Electronic Order Book Inevitable," *Journal of Finance*, 49, 1127–1161.

Goettler R. L., C. A. Parlour, and U. Rajan, 2003, "Equilibrium in a Dynamic Limit Order Market," forthcoming in *Journal of Finance*, 60, 2149–2192.

Goldstein M., and K. A. Kavajecz, 2000, "Eighths, Sixteenths, and Market Depth: Changes in Tick Size and Liquidity Provision on the NYSE," *Journal of Financial Economics*, 56, 125–149.

Griffith, M., B. Smith, D. Turnbull, and R. W. White, 2000, "The Costs and Determinants of Order Aggressiveness," *Journal of Financial Economics*, 56, 65–88.

Handa, P., and R. Schwartz, 1996, "Limit Order Trading, *Journal of Finance*," 51, 1835–1861.

Harris, L., 1998, "Optimal Dynamic Order Submission Strategies in Some Stylized Trading Problems," *Financial Markets, Institutions and Instruments*, 7, 1–91.

Harris, L., 2003, *Trading and Exchanges*, Oxford University Press, New York.

Harris, L., and J. Hasbrouck, 1996, “Market versus Limit orders: the Superdot evidence on Order Submission Strategy,” *Journal of Financial and Quantitative Analysis*, 31, 213–231.

Hasbrouck, J., 1999, *Trading Fast and Slow: Security Market Events in Real Time*, Mimeo, NYU.

Hasbrouck, J., and G. Saar, 2002, “Limit Orders and Volatility in a Hybrid Market: The Island ECN,” working paper, NYU.

Hollifield, B., A. Miller, and P. Sandås, 2004, “Empirical Analysis of Limit Order Markets,” *Review of Economic Studies*, 71, 1027–1063.

Hollifield, B., A. Miller, P. Sandås, and J. Slive, 2003, “*Estimating the Gains from Trade in Limit Order Markets*,” Mimeo, Carnegie Mellon University.

Huang, R., and H. Stoll, 1997, “The Components of the Bid-Ask Spread: A General Approach,” *Review of Financial Studies*, 10, 995–1034.

Jain, P., 2002, “Institutional Design and Liquidity on Stock Exchanges,” working paper, Indiana University.

Kadan, O., 2005, “So Who Gains From a Small Tick Size?,” forthcoming in *Journal of Financial Intermediation*.

Kaniel R., and H. Liu, 2005, “So What Orders Do Informed Traders Use?”, forthcoming in the Journal of Business.

Kavajecz, K., 1999, “A Specialist’s Quoted Depth and the Limit Order Book,” *Journal of Finance*, 54, 747–771.

Kavajecz, K., and E. Odders-White, 2003, “Technical Analysis and Liquidity Provision,” 17, 1043–1071.

Keim, D., and A. Madhavan, 1995, “Anatomy of the Trading Process: Empirical Evidence on the Behavior of Institutional Traders,” *Journal of Financial Economics*, 37, 371–398.

Kyle, R., 1985, “Continuous Auctions and Insider Trading,” *Econometrica*, 53, 1315–1335.

Lo, A., C. McKinlay, and J. Zhang, 2001, “Econometric Models of Limit Order Executions,” *Journal of Financial Economics*, 65, 31–71.

Madhavan, A., M. Richardson, and M. Roomans, 1997, “Why do Securities Prices Change? A Transaction-Level Analysis of NYSE Stocks,” *Review of Financial Studies*, 10, 1035–1064.

Pagano M., and R. Schwartz, 2003, “A Closing Call’s Impact on Market Quality at Euronext Paris,” *Journal of Financial Economics*, 68, 439–484.

Parlour, C., 1998, “Price Dynamics in Limit Order Markets,” *Review of Financial Studies*, 11, 789–816.

Parlour, C., and D. Seppi, 2003, “Liquidity-Based Competition for Order Flow,” *Review of Financial Studies*, 16, 301–343.

Rock, K., 1996, “The Specialist’s Order Book and Price Anomalies,” working paper, Harvard University.

Rosu, I., 2004, “A Dynamic Model of the Limit Order Book,” Mimeo, MIT.

Sandås, P., 2001, “Adverse Selection and Competitive Market Making: Evidence from a Limit Order Market,” *Review of Financial Studies*, 14, 705–734.

Seppi, D., 1997, “Liquidity Provision with Limit Orders and a Strategic Specialist,” *Review of Financial Studies*, 10, 103–150.

Tkach, I., 2002, “Liquidity Provision on the Tel Aviv Stock Exchange,” Mimeo, Hebrew University.

Tkach, I., and E. Kandel, 2004, “Demand for Immediacy – Time is Money,” Mimeo, Hebrew University.