A Dynamic Model of the Limit Order Book

Ioanid Roşu
University of Chicago

This paper presents a model of an order-driven market where fully strategic, symmetrically informed liquidity traders dynamically choose between limit and market orders, trading off execution price and waiting costs. In equilibrium, the bid and ask prices depend only on the numbers of buy and sell orders in the book. The model has a number of empirical predictions: (i) higher trading activity and higher trading competition cause smaller spreads and lower price impact; (ii) market orders lead to a temporary price impact larger than the permanent price impact, therefore to price overshooting; (iii) buy and sell orders can cluster away from the bid-ask spread, generating a hump-shaped order book; (iv) bid and ask prices display a comovement effect: after, e.g., a sell market order moves the bid price down, the ask price also falls, by a smaller amount, so the bid-ask spread widens; (v) when the order book is full, traders may submit quick, or fleeting, limit orders. (*JEL* C7, D4, G1)

1. Introduction

This paper presents a model of price formation in an order-driven market, where agents trade via a limit order book.[1] Compared with a quote-driven market, in which market makers provide liquidity by setting bid and ask quotes, in an order-driven market there are no designated market makers. Instead, liquidity is offered in a decentralized way, with anonymous traders who place orders in the limit order book, and wait until the orders get executed. Nowadays, more than half of the world's stock exchanges are order driven, with a limit order book at the center of the trading process (see Jain 2003).[2] Yet, despite

The author thanks Rob Battalio, Shane Corwin, Thierry Foucault, Drew Fudenberg, Xavier Gabaix, Larry Glosten, Burton Hollifield, Sergei Izmalkov, Eugene Kandel, Leonid Kogan, Jon Lewellen, Juhani Linnainmaa, Andrew Lo, David Musto, Stew Myers, Jun Pan, Christine Parlour, Anna Pavlova, Duane Seppi, Chester Spatt, Richard Stanton, Dimitri Vayanos, and Jiang Wang for helpful comments and suggestions. He is also grateful to participants at the NBER meeting, May 2004; WFA meeting, June 2005; and to seminar audiences at MIT, Berkeley, Notre Dame, Toronto, Kellogg, Carnegie Mellon, Michigan, Wharton, and Chicago. Send correspondence to Ioanid Roşu, Booth School of Business, University of Chicago, 5807 South Woodlawn Avenue, Chicago, IL 60637; telephone: 773-834-1826; fax: 773-834-0944. E-mail: irosu@uchicago.edu.

[1] The order book is the collection of all outstanding limit orders. Limit orders are price-contingent orders to buy (sell) if the price falls below (rises above) a prespecified price. A sell limit order is also called an *offer* (or *ask*), while a buy limit order is also called a *bid*. The lowest offer is called the *ask price*, or simply *ask*, and the highest bid is called the *bid price*, or simply *bid*.

[2] Examples of pure order-driven markets include Euronext, Helsinki, Hong Kong, Swiss, Tokyo, Toronto, and various electronic communication networks (Island, Instinet, Archipelago). There are also hybrid exchanges (NYSE, NASDAQ, London), in which market makers exist but have to compete with other traders, who supply liquidity by limit orders. In these markets, the number of transactions that involve a market maker is usually small (see Hasbrouck and Sofianos 1993).

doi:10.1093/rfs/hhp011 Advance Access publication April 2, 2009

their increasing importance, the literature on order-driven markets is relatively small.

One of the reasons for the scarcity of models of order-driven markets is the sheer complexity of the problem. Unlike in the case of a quote-driven market, where only one or perhaps a few market makers need to be modeled, a satisfactory model of an order-driven market should explain how prices arise from the interaction of a large number of anonymous traders, who arrive in the market at different times, choose whether to trade immediately or to wait, and can behave strategically by changing their orders at any time.

This paper presents a tractable dynamic model of an order-driven market that reflects the features mentioned above. To the author's knowledge, this is the first dynamic model in which agents are allowed to freely modify and cancel their limit orders. Surprisingly, allowing traders to be fully strategic turns out to simplify the problem, rather than complicate it. The model has a number of empirical implications about the bid-ask spread, trading volume, the price impact of transactions, and the evolution of the limit order book in time. Some of these implications provide a different interpretation of known empirical facts about order-driven markets, and some implications are new, and can be used to test the model.

It is interesting that all these implications are obtained in the absence of asymmetric information among traders. This is an advantage, because asymmetric information is hard to measure and is usually not observable. Instead, this paper relates the shape of the limit order book to observable quantities, such as the number of traders and their arrival rates.

Moreover, the paper offers a different interpretation of what determines the shape of the limit order book. In the market microstructure literature based on asymmetric information,[3] limit orders are placed at different levels because liquidity providers must protect themselves from traders with superior information. In particular, the bid-ask spread is smaller, and the limit order book is denser when there is less asymmetric information in the market. By contrast, in this paper, limit orders are placed on different levels because traders have to be compensated for their waiting costs. Since the compensation is determined by the spreads between limit orders, when there is more trading activity (i.e., traders arrive faster in the market), liquidity providers do not wait that much and so can be compensated with smaller spreads. Put differently, in this framework, a market is considered liquid if it is fast and/or competitive. In the market microstructure based on asymmetric information, a market is liquid if the amount of asymmetric information is small.

Specifically, this paper considers a continuous-time, infinite-horizon economy in which there is only one asset with no dividends. Buyers and sellers arrive in the market randomly. They either buy or sell one unit of the asset, after

[3] See, e.g., Glosten and Milgrom (1985); Kyle (1985); Easley and O'Hara (1987); and Glosten (1994). See O'Hara (1995) for a survey.

which they exit the model. It is assumed that all traders are liquidity traders, in the sense that their impulse to trade is exogenous to the model. However, they are discretionary liquidity traders in that they have a choice about when to trade and whether to place a market or limit order. After a limit order is placed, it can be canceled or changed at will. The execution of limit orders is subject to the usual price priority rule, and when prices are equal, to the time priority rule. All agents incur waiting costs—i.e., a loss of utility from waiting. Depending on whether they have low or high waiting costs, traders are patient or impatient. All information is common knowledge.

To find the equilibrium, one notices that all the traders on one side of the limit order book must have the same expected utility—otherwise, they would immediately undercut each other. And, because of the waiting costs, traders place their limit orders on different levels: for example, a seller with a limit order at a higher level gets a better expected price than a seller at a lower level, but has to wait longer—in such a way that both sellers get the same expected utility. This implies that there is a (Markov perfect) equilibrium where only the number of buyers and sellers matters. Therefore, their expected utility follows a recursive system of difference equations. The recursive system can be solved numerically, and in some cases in closed form.

In equilibrium, impatient agents submit market orders, while patient agents submit limit orders and wait, except for the states in which the limit order book is full.[4] In states in which the book is not full, new limit orders are always placed inside the bid-ask spread.[5] The point where the book is full coincides with the time when the bid-ask spread is at the minimum. That there exists a nonzero minimum bid-ask spread is an interesting fact since the tick size is zero in this model.

Particular cases of the model can be solved in closed form. One important example is when there are only patient sellers and impatient buyers (or buyers can place only market orders). Section 3 shows how to use these formulas to derive implications about the average bid-ask spread and price impact and the average maximum number of traders in the book, and how to use the model to estimate in practice the arrival rates of patient and impatient traders. For example, in markets with higher trading activity (measured by the sum of arrival rates of all agents) and with higher competition (measured by the ratio of arrival rates for patient and impatient traders) both the bid-ask spreads and the price impact will be smaller.[6] An intriguing implication is that more trading

[4] When the book is full, some patient agent either places a market order or submits a quick (fleeting) limit order, which some trader from the other side of the book immediately accepts. This comes theoretically as a result of a game of attrition among the buyers and sellers.

[5] In their analysis of the Paris Bourse market (now Euronext), Biais, Hillion, and Spatt (1995) observe that the majority of limit orders are spread improving.

[6] These predictions are tested for the Helsinki Exchange by Linnainmaa and Roşu (2008), using weather as an instrument. Trading activity has been shown to explain variation in spreads, starting with Demsetz (1968). Some evidence that reasons other than information may better explain prices can also be found, e.g., in Huang and

activity leads to lower asset volatility; more precisely, the volatility of the asset must vary in inverse proportion to the square of the trading activity.

In order to discuss price impact and determine the distribution of limit orders in the order book, the model next allows multi-unit market orders to arrive with positive probability. Then one can define both the *temporary* (or instantaneous) price impact function, which is the actual price impact suffered by the market order trader, and the *permanent* (or subsequent) price impact, which is the difference between the new ask price and the ask price before the market order was submitted. In this setup, the temporary price impact is larger than the permanent price impact, which is equivalent to price overshooting. The intuition is that, before a multi-unit market order comes, the traders who expect their limit orders to be executed do not know the exact order size, so they stay higher in the book. Once the size becomes known, the sellers regroup lower in the book.

Also, if multi-unit market orders arrive with probabilities that do not decrease too fast with order size, then the price impact function is typically first concave and then convex. This is the same as saying that the limit orders cluster away from the bid and the ask, or that the book exhibits a "hump" shape.[7] This arises because patient traders cluster away from the bid-ask spread when they expect to take advantage of large market orders that are not too unlikely.

The general, two-sided case is more difficult, and the solution must be found numerically. An empirical implication is the *comovement* effect between bid and ask prices.[8] For example, a market sell order not only decreases the bid price—due to the mechanical execution of limit orders on the buy side—but also subsequently decreases the ask price. Moreover, the decrease in the bid price is larger than the subsequent decrease in the ask price, which leads to a wider bid-ask spread. The comovement effect is stronger when there are more limit traders on the side of the subsequent price move, and the competition among them is stronger.

Another consequence of the general case is the existence of quick (or fleeting) limit orders: when the limit order book becomes full, a buyer or seller places a limit order, and a limit trader on the other side immediately accepts it by canceling the limit order and placing a market order. In this model, fleeting limit orders appear only when the order book is full, and are always placed between the bid and the ask.

The comovement effect and the presence of temporary and permanent price impact raise the issue of comparing the predictions of the current model based on waiting costs with the predictions of a model based on asymmetric

---

Stoll (1997), who estimate that on average approximately 90% of the bid-ask spread is due to noninformational frictions ("order-processing costs").

[7] See Bouchaud, Mezard, and Potters (2002, Figure 2); or Biais, Hillion, and Spatt (1995, p. 1664 ff.).

[8] Biais, Hillion, and Spatt (1995) document a comovement effect in the Paris Bourse (now Euronext), and suggest an explanation based on asymmetric information.

information. Section 6 suggests one way of telling these two stories apart, based on the behavior of the bid-ask spread.

The limit order book has been analyzed in a variety of ways. The static models are typically based on asymmetric information: see Glosten (1994); Chakravarty and Holden (1995); Handa and Schwartz (1996); Rock (1996); and Seppi (1997). The present paper is part of a more recent literature that analyzes dynamic aspects of the limit order book, usually in the absence of asymmetric information. A precursor of this literature is the "gravitational pull" model of Cohen et al. (1981), in which traders choose between limit and market orders based on their expectations about the evolution of an exogenous price process. Parlour (1998) proposes a two-tick model where traders choose between limit orders and market orders after taking into account the effect of their decision on future traders' strategies; the model can explain various patterns in order placement strategies, including the comovement effect. Foucault (1999) studies the choice between limit and market orders and focuses on the nonexecution probability of limit orders and the winner's curse problem for the limit orders when they do execute. Goettler, Parlour, and Rajan (2005) solve numerically for the stationary Markov perfect equilibrium in a model in which traders with private valuations choose whether to submit a market or a limit order, and also choose the size of the order. The paper most closely related to the present one is that of Foucault, Kadan, and Kandel (2005), in which traders also face waiting costs, but cannot modify their orders once submitted; they focus their analysis on the bid-ask spread, market resiliency, and time to execution for limit orders.

The paper is organized as follows. Section 2 describes the model. Section 3 solves for the equilibrium in a particular case that represents the sell side of the book: there are only patient sellers and impatient buyers. Section 4 discusses the case of multi-unit market orders and analyzes the price impact. Section 5 describes the equilibrium in the general case with all types of sellers and buyers and derives implications about the comovement effect and fleeting orders. Section 6 contrasts the present model based on waiting costs to a model based on information. Section 7 concludes.

2. The Model

2.1 The market

This section describes the assumptions of the model. Consider a market for an asset that pays no dividends. The buy and sell prices for this asset are determined as the bid and ask prices resulting from trading based on the rules given below. There is a constant range $A > B$ in which the prices lie at all times.[9] More specifically, there is an infinite supply when price is A, provided

[9] One can think about A and B as summarizing the information about the asset: $(A + B)/2$ represents the average value of the asset, while $(A - B)$ represents differences of opinion among traders. See Roşu (2008) for a theoretical model in which the bounds A and B vary over time along with the efficient price, and their difference depends on agents' private costs of trading.

by agents outside the model. Similarly, there is an infinite demand for the asset when price is B. Prices can take any value in this range—i.e., the tick size is zero.

Trading. The time horizon is infinite, and trading in the asset takes place in continuous time. The only types of trades allowed are market orders and limit orders. The limit orders are subject to the usual price priority rule, and, when prices are equal, the time priority rule is applied. If several market orders are submitted at the same time, only one of them is executed, at random, while the other orders are canceled.[10]

Limit orders can be canceled for no cost at any time.[11] There is also no delay in trading, both types of orders being posted or executed instantaneously. Trading is based on a publicly observable limit order book.

Agents. Traders arrive randomly in the market, due to liquidity needs that are not modeled here. The arrival process is assumed to be exogenous and is described in more detail below. Once traders arrive, they choose strategically between market and limit orders. The traders are either buyers or sellers; their type is fixed from the beginning and cannot change. Buyers and sellers trade at most one unit, after which they exit the model forever.[12] Traders are risk neutral, so their instantaneous utility function equals the price for sellers, minus the price for buyers. As in Foucault, Kadan, and Kandel (2005), traders lose utility proportionally to their expected waiting time.[13] If τ is the random execution time and P_τ is the price obtained at τ, the expected utility of a seller with patience coefficient $\tilde{r}$ is $f_t = \mathsf{E}_t\{P_\tau - \tilde{r}(\tau - t)\}$. (The expectation operator takes as given the strategies of all the players. See the description of the strategies below.) Similarly, the expected utility of a buyer is $-g_t = \mathsf{E}_t\{-P_\tau - \tilde{r}(\tau - t)\}$, where we denote $g_t = \mathsf{E}_t\{P_\tau + \tilde{r}(\tau - t)\}$. Notice that g_t equals *minus* the expected utility of a buyer; this is done in order to compare buyers and sellers more easily.

The discount coefficient $\tilde{r}$ is a constant that can take only two values $r < r'$. If $\tilde{r} = r$, the trader is called *patient*; and if $\tilde{r} = r'$, the trader is called *impatient*.

---

[10] To justify this assumption, think of a market buy/sell order as a (marketable) limit order at a price equal to the ask/bid. Then, if several market orders are submitted at the same time, one of them is randomly executed, while the others remain as limit orders, which can be freely canceled.

[11] In most financial markets, cancellation of a limit order is free, although one may argue that there are still *monitoring* costs. Foucault, Kadan, and Kandel (2005) consider a model with infinite monitoring costs, in which agents never change their orders once submitted.

[12] This is a strong assumption: some agents might want to trade larger quantities or decide to remain in the market to buy and sell securities, thus in effect becoming market makers. (Bloomfield, O'Hara, and Saar 2005 show experimentally that market making arises endogenously in pure limit order markets.) But, as long as liquidity suppliers have some constraints due to inventory reasons or risk aversion, one could adopt a model similar to the present one but replace the one-unit limit with an n-unit limit.

[13] The paper is deliberately vague about the exact nature of the waiting costs. Besides the standard time discounting story, one can also think of the opportunity cost of trading in the given asset. Furthermore, one can interpret the waiting costs as uncertainty aversion, which increases with the time horizon.

For simplicity, it is assumed that r' is much larger than r, which implies that impatient traders always submit market orders.[14]

Arrivals. The four types of traders (patient buyers, patient sellers, impatient buyers, and impatient sellers) arrive in the market according to independent Poisson processes with constant, exogenous intensity rates $\lambda_{PB}, \lambda_{PS}, \lambda_{IB}, \lambda_{IS}$.[15] By definition, a Poisson arrival with intensity λ implies that the number of arrivals in any interval of length T has a Poisson distribution with parameter λT. The inter-arrival times of a Poisson process are distributed as an exponential variable with the same parameter λ. The mean time until the next arrival is then $1/\lambda$.

In the rest of the paper, to say that an event happens after Poisson (ν) means that the event time coincides with the first arrival in a Poisson process with intensity ν.

Strategies. Since this is a model of continuous trading, it is desirable to set the game in continuous time. There are also technical reasons why that would be useful: in continuous time, with Poisson arrivals the probability that two agents arrive at the same time is zero. This simplifies the analysis of the game.

Another important benefit of setting the game in continuous time is that agents can respond immediately. More precisely, one can use strategies that specify: "Keep the limit order at a_1 as long as the other agent stays at a_2 or below. If at some time t the other agent places an order above a_2, then *immediately after* t undercut at a_2." Immediate punishment allows simple solutions, whereby existing traders do not need to change their strategy until the arrival of the next trader.

Setting the game in continuous time, nevertheless, requires extra care. We use the framework developed in Roşu (2006), which allows for multi-stage games and mixed strategies.[16] The types of equilibrium used are subgame perfect

---

[14] This can be shown using Proposition 12 in the Appendix: in a game of attrition with both patient and impatient traders, a significantly more impatient trader has no reason to wait.

[15] Biais, Hillion, and Spatt (1995) show empirically that arrivals are positively correlated: the "diagonal effect"; and that they depend on the state of the order book: the larger the bid-ask spread, the faster limit orders arrive to supply liquidity (Hollifield et al. 2006 argue that the opposite is true on the Vancouver stock exchange). See Roşu (2008) for a theoretical model that endogenizes entry decisions, and produces positively correlated arrivals that depend on the state of the order book.

[16] The main difficulty in continuous time game theory comes from the fact that given a time t, there is no last time before t and no first time after t. The solution is to allow strategies with infinitesimal inertia (as in Bergin and MacLeod 1993) and a uniformly bounded number of jumps (as in Simon and Stinchcombe 1989). Infinitesimal inertia means that agents do not change their strategies in the infinitesimal time interval $[t, t + \mathrm{d}t]$ (this is similar to continuous time finance, where agents do not trade during $[t, t + \mathrm{d}t]$). The extension to multistage game theory is needed because of market orders: when a market order arrives at time t, an existing limit trader exits the model, and the next stage of the game must take place with fewer traders, at the same time t. This can be done by "stopping the clock," so that the next game is also played at t. Also, in continuous time there can be both *mixing over actions,* by choosing randomly an action in the stage game at time t; and *mixing over time,* by starting with a deterministic choice at time t but changing the action randomly in the interval $(t, t + \mathrm{d}t)$. Since in this paper nature mixes over time by bringing agents according to a Poisson process, it is most natural to consider strategies mixed over time.

equilibrium, and Markov perfect equilibrium (see Fudenberg and Tirole 1991, chap. 13). Another important notion in this framework is that of *competitive* Markov equilibrium, which is a Markov perfect equilibrium from which local deviations can be stopped by local punishments—assuming that the behavior in the rest of the game does not change. In other words, if a local deviation improves a trader's expected payoff—ignoring what happens in the other states of the game—the trader would deviate.

Also, for the purposes of this paper, one introduces the notion of *rigid* equilibrium, which is a competitive stationary Markov equilibrium in which, if some agents have mixed strategies, mixing is done only by the agents with the most competitive limit orders (highest bid or lowest offer).[17]

Finally, in this paper all information, including agents' strategies and beliefs, is common knowledge.

3. Equilibrium: One Side of the Book

This section analyzes the particular case in which all the sellers are patient and all the buyers are impatient. (By symmetry, one can derive similar results for the market with only patient buyers and impatient sellers.) So assume that the arrival rates of patient buyers and impatient sellers are zero: $\lambda_{PB} = \lambda_{IS} = 0$. To simplify notation, denote the arrival rates of patient sellers and impatient buyers, respectively, by

$$\lambda_1 = \lambda_{PS} \quad \text{and} \quad \lambda_2 = \lambda_{IB}.$$

This case is very tractable and has closed-form solutions. Moreover, while potentially useful in its own right, it is very important in getting intuition about the general case. Indeed, one can regard the sell side of a general limit order book as being driven only by patient sellers and impatient buyers, except that the lower bound B of the sell side is not fixed, but equals the bid price—i.e., the highest bid on the buy side.

3.1 Main intuition

Suppose that the limit order book is empty, and a patient seller labeled "1" arrives first in the market. Then trader 1 submits a limit sell order at the maximum level $a_1 = A$ and remains a monopolist until some other trader arrives.[18] Suppose that a second patient seller labeled "2" arrives. Now both sellers compete for market orders from the incoming impatient buyers. If trader 1 could not cancel his limit order at A, then trader 2 would undercut by placing

[17] In the language of Corollary 2 in the Appendix, in case 4 only equilibria of type c occur.

[18] It is assumed implicitly that if the only limit sell orders in the book are at A, a market order first clears the orders in the book, and only then relies on the infinite supply at A.

a limit order at $a_2 = A - \delta$ for some very small δ.[19] Her expected utility would then be strictly larger than that of trader 1. But trader 1 can change his limit order, so a price war would follow. Undercutting happens instantaneously in this model, because the game is set in continuous time (see the discussion about strategies in Section 2).

As a result, trader 1 does not need to change his limit order as long as trader 2 places her limit order at some level $a_2 < a_1 = A$ that is low enough. To find the correct level for a_2, note that both traders must have the same expected utility in equilibrium. If trader 2 placed her order above a_2 where she had higher expected utility than trader 1, then trader 1 would immediately undercut by a penny, and so on. So, in the equilibrium with two sellers, trader 1 has a limit order at $a_1 = A$, and trader 2 has a limit order at $a_2 < A$. Both traders have the same expected utility: trader 1 obtains in expectation a higher price than trader 1, but waits longer to get his order executed.[20] This leads to an equilibrium in which patient sellers compete with offers at different prices in order to extract rents from the impatient buyers.[21]

In solving for the equilibrium, it is surprising that allowing agents to freely cancel or modify their limit orders, instead of complicating the solution, actually simplifies it. This happens because the threat of undercutting makes all patient sellers have the same expected utility in equilibrium. This makes the equilibrium Markov, with the number of sellers as a state variable.

An important property of this equilibrium is that it is competitive, in the sense that a local deviation from one of the traders can be stopped by another trader's immediate undercutting, assuming that the rest of the equilibrium behavior does not change. One can also imagine a noncompetitive equilibrium. For example, suppose that all patient sellers queue their limit orders at A until the expected utility of the last trader equals the reservation value B. This equilibrium is sustained by Nash threats: trader 1 threatens with competitive behavior if trader 2 does not queue behind him at A. Trader 2 is better off complying as long as she expects trader 3 to do the same and queue behind her. This equilibrium is noncompetitive because punishment implies that behavior in the rest of the game will be changed (to the competitive equilibrium). Noncompetitive

---

[19] In Foucault, Kadan, and Kandel (2005), even though trader 1 cannot cancel his limit order, trader 2 would still undercut by more than a penny in equilibrium. This is because the strategy of future arriving buyers depends on the level of trader 2's limit order: the higher it is, the less likely it is that a buyer will place a market order. Therefore, in their model it is important that traders on the other side of the limit order book are able to place limit orders. In this model, the main intuition is robust regardless whether it is a one-sided or a two-sided limit order book.

[20] The intuition is supported by empirical work. Lo, Mackinlay, and Zhang (2002) document that execution times are very sensitive to the limit price (but not to the limit order size). See also Hollifield, Miller, and Sandas (2004).

[21] Unlike the Bertrand price competition, in this model traders make positive expected profits: waiting costs generate an equilibrium with offers at different prices, leading to an expected utility above their reservation value B. This is consistent with the empirical literature, which shows that limit order traders make positive profits: see Sandås (2001) for the case of Stockholm Stock Exchange and Harris and Hasbrouck (1996) for NYSE's SuperDOT system. See also Biais, Martimort, and Rochet (2000) for a theoretical model in which market makers competing with supply schedules make positive profits.

equilibria may be important, for example, in understanding dealer markets.[22] The present paper focuses on competitive equilibria, since they are the more likely outcome of large, anonymous order-driven markets.

3.2 Description of the equilibrium

Consider a limit order book with upper bound A; the lower bound B (which is the reservation value of the sellers, as they can always submit a market order for B and exit the model); the sellers' patience coefficient r; the patient seller arrival rate λ_1; and the impatient buyer arrival rate λ_2.

Before a more formal discussion of the results, some intuition is given about how the equilibrium works (proofs will be given later). As discussed above, in equilibrium, all patient sellers in the order book have the same expected utility. Denote the number of sellers by m, and their expected utility by f_m. This means that, in a Markov equilibrium, the number of sellers m is a *state variable*. The number of m evolves according to a Markov process, and so the sellers' utility f_m satisfies a system of equations, called the *recursive system.*

The number of states must be finite: otherwise, the expected execution time of the top-limit seller would be infinite, hence his utility would be negative infinity; but then he would rather submit a market order at the lower bound B, which is the reservation value of the sellers. As the number of sellers m increases, each seller is strictly worse off, and the ask price will decrease. Denote by M the largest number of limit orders the equilibrium book can accommodate. A limit order book with the maximum number of orders M is called "full." In that case, it must be that the expected utility of each of the existing M sellers equals $f_M = B$: otherwise, if $f_M > B$ an incoming patient seller would want to join in to get more than the reservation value B. Now, if the sellers' utility f_M *exactly* equals the reservation value B, it must be that one of the sellers (by choice the bottom seller—i.e., the one with the lowest offer) has a mixed strategy: after Poisson(ν) the bottom seller places a market order at B and exits.

Observe that from a limit order book with m sellers ($m = 1, \ldots, M-1$), the market can go either to state $m+1$ if a patient seller arrives—after random time $T_1 \sim \exp(\lambda_1)$; or to state $m-1$ if an impatient buyer arrives—after random time $T_2 \sim \exp(\lambda_2)$. Inter-arrival times of Poisson processes are exponentially distributed, so the arrival of the first of the two states happens at $T = \min(T_1, T_2)$, which is exponential with intensity $\lambda_1 + \lambda_2$ (hence the expected value of T is $\frac{1}{\lambda_1+\lambda_2}$). The first event happens with probability $\frac{\lambda_1}{\lambda_1+\lambda_2}$, while the second event happens with probability $\frac{\lambda_2}{\lambda_1+\lambda_2}$. One obtains the formula

$$f_m \;=\; \frac{\lambda_1}{\lambda_1+\lambda_2} f_{m+1} \;+\; \frac{\lambda_2}{\lambda_1+\lambda_2} f_{m-1} \;-\; r \cdot \frac{1}{\lambda_1+\lambda_2}.$$

[22] Christie and Schultz (1994) document collusion among NASDAQ dealers in the early 1990s. Their paper contributed to the 1997 introduction in NASDAQ of a public limit order book.

Notice also that there are two types of sellers: the bottom seller, who has a limit sell order at the ask price a_m, and the other sellers, who have their limit orders above a_m. If an impatient buyer arrives and places a market order, the bottom seller receives a_m, while the other sellers get f_{m-1}, which is the expected utility of sellers in an order book with $m-1$ sellers. Since all sellers must have the same expected utility in equilibrium, it must be that the ask price a_m equals the expected utility where there is one less seller:

$$a_m \;=\; f_{m-1}.$$

From the state with the maximum number of sellers M, the system can go only to state $M-1$, either if an impatient buyer arrives—after random time $T_1 \sim \exp(\lambda_2)$ or if the seller with the current bottom limit order places a market order at B and exits—after random time $T_2 \sim \exp(\nu)$.[23] Then one obtains the formula

$$f_M \;=\; f_{M-1} \;-\; r \cdot \frac{1}{\lambda_2 + \nu}.$$

Define also $f_0 = A$.[24] In conclusion, f_m satisfies a system of difference equations, called the recursive system.

Definition 1. *Start with the limit order book upper bound A and lower bound B, the arrival rates of patient sellers λ_1 and impatient buyers λ_2, and the sellers' patience coefficient r. Then the* recursive system *is a collection (f_m, M, ν) of the sellers' expected utility f_m, the maximum number of sellers in the book $M > 0$, and the mixed strategy Poisson rate for the bottom seller $\nu \geq 0$, which satisfy*

$$\begin{cases} f_0 = A, \\ f_m = \dfrac{\lambda_1}{\lambda_1+\lambda_2} f_{m+1} + \dfrac{\lambda_2}{\lambda_1+\lambda_2} f_{m-1} - r \cdot \dfrac{1}{\lambda_1+\lambda_2}, \quad m = 1, \ldots, M-1, \\ f_M \;=\; f_{M-1} \;-\; r \cdot \dfrac{1}{\lambda_2+\nu}, \\ f_M = B. \end{cases} \tag{1}$$

The next theorem describes the equilibrium order book resulting from the solution to the recursive system. Recall that a rigid equilibrium is a competitive stationary Markov perfect equilibrium in which, if some agents have mixed

[23] One can ignore the arrival of a new patient seller, because, in equilibrium, he will immediately place a market order at B and exit, without affecting the state.

[24] This is justified by Theorem 1, in the proof of which one can see that $a_m = f_{m-1}$ for all $1 < m < M$. Since the sole trader at $m = 1$ places an order at $a_1 = A$, one has by extension $A = a_1 = f_0$.

strategies, mixing is done only by the agents with the most competitive limit orders (in this case, only by the seller with the lowest offer).

Theorem 1. *Consider the limit order book with only patient sellers and impatient buyers ($\lambda_{PB} = \lambda_{IS} = 0$), with upper bound A, lower bound B, the arrival rates of patient sellers λ_1 and impatient buyers λ_2, and the sellers' patience coefficient r. Then there exists a Markov perfect equilibrium of the game, with a maximum number M of sellers, such that in the state with $m \leq M$ patient sellers, the sellers' expected utility f_m is given by*

$$f_m = A + C\left(\left(\frac{\lambda_2}{\lambda_1}\right)^m - 1\right) + \frac{r}{\lambda_1 - \lambda_2} m, \quad \text{if } \lambda_1 \neq \lambda_2, \tag{2}$$

$$f_m = A - bm + \frac{r}{\lambda_1 + \lambda_2} m^2, \quad \text{if } \lambda_1 = \lambda_2, \tag{3}$$

where $C > 0$, $b > 0$ and $M > 0$ are defined in Proposition 11 in the Appendix. The ask price a_m in the state with m sellers is given by

$$a_m = f_{m-1}, \quad \text{if } m < M; \tag{4}$$

$$a_M = B + \frac{r}{\lambda_2}. \tag{5}$$

The strategy of each agent in the state with m sellers is the following: If $m = 1$, then place a limit order at $a_1 = A$. If $m = 2, \ldots, M-1$, place a limit order at any level above a_m, as long as someone has stayed at a_m or below; otherwise place an order at a_m. If $m = M$, the strategy is the same as for $m = 2, \ldots, M-1$, except for the bottom seller at a_M, who exits at the first arrival in a Poisson process with rate $\nu \geq 0$ by placing a market order at B (the number $\nu \geq 0$ is defined in Proposition 11 in the Appendix). If $m > M$, then immediately place a market order at B.

The equilibrium described above is Markov, with state variables: the number of existing sellers and the ask price.[25] *This equilibrium is unique in the class of rigid equilibria, in the sense that any other rigid equilibrium leads to the same evolution of the state variables.*

Proof. See the Appendix. ■

It should be pointed out that there is some ambiguity in the way strategies are formulated: in the state with m sellers, as long as some seller has a limit order at the ask price a_m (or below), the other sellers can place their limit orders anywhere above a_m. From now on, by an abuse of notation, all the equilibria in this class are considered to be the same equilibrium. Moreover,

[25] Since the equilibrium ask a_m is a function of m, it may seem that m is the only state variable in this Markov equilibrium. In fact, the ask price is also a state variable, since it describes what happens out of equilibrium: if the seller at the ask a_m suddenly increases his order, then some other seller would immediately lower her order exactly to the level a_m.

one can choose in each class a particularly important representative, called the canonical equilibrium.

Definition 2. *To define the* canonical equilibrium*, suppose that a new seller arrives when there are already $m-1$ sellers in the book. Then the equilibrium strategies require that the new seller place an order at a_m, while the others stay on their previous levels. The outcome of this equilibrium is that, in state m, sellers have their offers placed at $a_1, \ldots a_m$, and they never change them.*

The canonical equilibrium arises naturally if one introduces an infinitesimal cancellation cost for limit orders: sellers have no reason to modify their limit orders if it is costly to do so. The canonical equilibrium also appears as a limiting case of the equilibrium when multi-unit market orders are allowed but very unlikely (see Proposition 6 in Section 4).

Now we move to deriving a few implications of Theorem 1 that will be useful later. The first one shows that when the limit order book is full (in the state M with the maximum number of sellers), there is a minimum bid-ask spread. This is interesting, given that prices can take any value (i.e., the tick size is zero).

Corollary 1. *There exists a minimum bid-ask spread $S_{\min}$ given by*

$$S_{\min} = a_M - B = \frac{r}{\lambda_2}. \tag{6}$$

One can also compute the dependence of the sellers' expected utility f_m on the lower bound B of the order book. This is important for the general (two-sided) limit order book, since there the sell side can be regarded as a one-sided order book in which the lower bound B is not constant but equals the bid price.

Proposition 1. *If patient sellers and impatient buyers arrive at different rates (i.e., $\lambda_1 \neq \lambda_2$), then the expected utility of m sellers f_m depends on lower bound B in the following way: $\mathrm{d}f_m/\mathrm{d}B = 1 - (\frac{\lambda_2}{\lambda_1})^m / 1 - (\frac{\lambda_2}{\lambda_1})^M$. The derivative is computed by holding the number of states M constant.*

Proof. Denote $\alpha = \frac{\lambda_2}{\lambda_1}$. Differentiate the formula $f_m = A + C(\alpha^m - 1) + \frac{r}{\lambda_1 - \lambda_2} m$ with respect to B: $\frac{\mathrm{d}f_m}{\mathrm{d}B} = \frac{\mathrm{d}C}{\mathrm{d}B}(\alpha^m - 1)$. For $m = M$ the formula for f_m becomes $B = A + C(\alpha^M - 1) + \frac{r}{\lambda_1 - \lambda_2} M$. Differentiating this with respect to B and holding M constant, one gets $1 = \frac{\mathrm{d}C}{\mathrm{d}B}(\alpha^M - 1)$, so $\frac{\mathrm{d}C}{\mathrm{d}B} = \frac{1}{\alpha^M - 1}$. Finally, $\frac{\mathrm{d}f_m}{\mathrm{d}B} = \frac{1-\alpha^m}{1-\alpha^M}$. ∎

3.3 Empirical implications

Having described a closed-form solution for the one-sided limit order book (with only patient sellers and impatient buyers), we now derive empirical implications about the mean and standard deviation of the bid-ask spread and

price impact, and for the maximum possible number of limit orders in the book M. Some implications provide a different interpretation of known empirical facts about order-driven markets, and some implications are new.

Given the arrival rate of patient sellers λ_1, their patience coefficient r, and the arrival rate of impatient buyers λ_2, define three more numbers: the *activity* parameter λ: the arrival rate of all types of agents; the *competition* parameter c: the ratio of arrival rates of patient to impatient traders;[26] and the *granularity* parameter ε: the ratio of the sellers' patience coefficient to trading activity:[27]

$$\lambda = \lambda_1 + \lambda_2 \ = \ \textit{activity}, \tag{7}$$

$$c = \frac{\lambda_1}{\lambda_2} \ = \ \textit{competition}, \tag{8}$$

$$\varepsilon = \frac{r}{\lambda} \ = \ \textit{granularity}. \tag{9}$$

For now it is assumed that both trading activity λ and competition c are directly observable. Later in this section, it is shown how to use the model to estimate them.

It turns out that the limit order book behaves very differently depending on the competition parameter c. For example, when the patient sellers arrive faster than the impatient buyers ($c > 1$), the limit order book is *resilient*—i.e., the bid-ask spread on average reverts to smaller values. By contrast, if the patient sellers arrive at the same speed as the impatient buyers ($c = 1$) or slower ($c < 1$), the bid-ask spread can be quite wide (of the order of $A - B$, the difference between the upper bound and lower bound of the book). This is unreasonable, except perhaps in the case of very illiquid markets.

For the rest of this section, only the case $c > 1$ will be used. The next result gives the dependence of the mean and standard deviation of the bid-ask spread on the granularity ε and competition c.[28] Since the exact formulas are more complicated, it is more instructive to report the approximate formulas[29] only when trading activity λ is large (or equivalently, the granularity $\varepsilon = r/\lambda$ is small).

Proposition 2. *Consider a limit order book with only patient sellers and impatient buyers, with upper bound A, lower bound B, sellers' patience*

---

[26] The competition parameter c is also present in Foucault, Kadan, and Kandel (2005) (where it is called ρ) and is argued to be a key determinant of the *resilience* of the bid-ask spread—i.e., the tendency of the spread to return to lower levels.

[27] The term "granularity" is borrowed from the econophysics literature (see Farmer, Patelli, and Zovko 2005), where it is related to the size of gaps in the limit order book. Similarly, in our case, the granularity ε is of the order of the minimum bid-ask spread, as seen from Corollary 1.

[28] The corresponding formulas when $c \leq 1$ are in the Appendix, in the proof of Proposition 2.

[29] To be precise about what "approximate" means in this context, one says that a variable X is asymptotically equal to $f(\varepsilon)$ and writes $X \approx f(\varepsilon)$ if X can be written as $X = f(\varepsilon) + g(\varepsilon)$, with $\lim_{\varepsilon \to 0} \frac{g(\varepsilon)}{f(\varepsilon)} = 0$ (in standard mathematical notation, this is written as $X = f(\varepsilon) + o(f(\varepsilon))$).

coefficient r, trading activity λ, and competition c. Assume that sellers arrive faster than buyers—i.e., $c > 1$. Let $S_m = a_m - B$ be the equilibrium bid-ask spread in the state with m sellers. Then, when granularity $\varepsilon = \frac{r}{\lambda}$ is small, the mean spread $\bar{S}$ and the standard deviation $\sigma(S)$ can be approximated by

$$\bar{S} \approx \varepsilon \ln(1/\varepsilon)\, \frac{c(c+1)}{(c-1)\ln(c)}, \quad \sigma(S) \approx \sqrt{\varepsilon(A-B)}\, \left(\frac{(c+1)(c^3+c^2-c)}{(c-1)^3} \right)^{1/2}. \tag{10}$$

Both $\bar{S}$ and $\sigma(S)$ decrease with c as long as $c < 5$, and increase with ε; therefore, $\bar{S}$ and $\sigma(S)$ increase with sellers' patience r and decrease with trading activity λ.

Proof. See the Appendix. ■

Notice that the average bid-ask spread is of the order of $\varepsilon \ln(1/\varepsilon)$, where $\varepsilon = r/\lambda$ is the granularity parameter. This should be compared with Farmer, Patelli, and Zovko (2005), who in their cross-sectional empirical analysis of the London Stock Exchange show that with a high R^2 the average bid-ask spread varies proportionally to $\varepsilon^{3/4}$, which is close to the theoretical term $\varepsilon \ln(1/\varepsilon)$. It should be mentioned that their granularity parameter does not include the patience coefficient r, which is not directly observable.

An important implication of Proposition 2 is that the average bid-ask spread $\bar{S}$ is smaller when (i) sellers are more patient (r is smaller); (ii) traders arrive faster in the market (trading activity λ is higher).[30] Trading activity has been known to explain variation in spreads (see, e.g., Demsetz 1968), but it is important to point out that in this model there is a *causal* connection from trading activity to spreads. Linnainmaa and Roşu (2008) test this causal connection for the Helsinki Exchange by using weather in Finland as an instrument, and find that, indeed, higher trading activity causes smaller spreads.

Proposition 2 also implies that the average bid-ask spread is smaller when competition $c = \lambda_1/\lambda_2$ is higher—i.e., when patient sellers arrive at a faster rate relative to the impatient buyers. This result reverses when c is very high. This comes from the fact that the average bid-ask spread responds differently to sellers' and buyers' arrival rates λ_1 and λ_2. Competition c is higher either when λ_1 is higher or when λ_2 is lower. When patient sellers arrive more quickly (λ_1 is higher), the bid-ask spread is indeed smaller, due to increased competition among patient sellers. But when impatient buyers arrive more slowly (λ_2 is lower), the bid-ask spread is actually larger, indicating that patient sellers need a higher spread as compensation for waiting more. Normally, the first effect dominates, but when competition c is very high ($c > 5$), the second effect dominates.

[30] These results are also true in the model of Foucault, Kadan, and Kandel (2005).

Another implication about spreads can be derived from Corollary 1 in Section 3.2:

$$S_{\min} \;=\; \frac{r}{\lambda_2} \;=\; \frac{r}{\lambda}\,(c+1) \;=\; \varepsilon\,(c+1). \tag{11}$$

Notice that, unlike the average bid-ask spread $\bar{S}$, when competition c is higher, the minimum spread $S_{\min}$ is also *higher*.[31] This is because the minimum bid-ask spread depends only on the arrival rate of impatient buyers λ_2, and as before spreads are larger when impatient buyers arrive more slowly.

For the next result, one identifies the volatility of the asset as the standard deviation of the ask price (since the bid price B is constant), or equivalently the volatility of the spread $\sigma(S)$.

Proposition 3. *In the context of Proposition 2, the volatility of the asset $\sigma(S)$ varies in inverse proportion to $\sqrt{\lambda}$, the square root of trading activity. Also, the average spread $\bar{S}$ varies approximately proportionally to the ratio $\sigma(S)/\sqrt{\lambda}$.*

Proof. Equation (10) from Proposition 2 implies that the volatility $\sigma(S)$ varies in proportion to the square root of granularity $\varepsilon = r/\lambda$—i.e., in inverse proportion to $\sqrt{\lambda}$.

For the second result, use again Equation (10) to compute the ratio $\bar{S}/\sigma(S)$. This is proportional to $\sqrt{\varepsilon}\ln(1/\varepsilon)$. If one omits the term $\ln(1/\varepsilon)$, which goes to infinity at a much slower rate than the term $\sqrt{\varepsilon}$ goes to zero, then indeed $\bar{S}/\sigma(S)$ is proportional to $\sqrt{\varepsilon}$, and hence varies in inverse proportion to $\sqrt{\lambda}$. ■

Empirical evidence for both results can be found in Wyart et al. (2008, Equations (32) and (33)), for high-frequency data. Interestingly, the first result is in contradiction with a large literature that documents a positive relation between trading volume and volatility when dealing with lower-frequency data (see, e.g., Jones, Kaul, and Lipson 1994). The explanation is that Proposition 3 is a high-frequency result: it assumes the bounds A and B of the limit order book to be constant, which is reasonable only for intra-day time intervals.

Next, we define price impact in this context. Consider the canonical equilibrium of Definition 2. For simplicity, price impact is defined only for one-unit market orders, leaving the case of multi-unit market orders for the next section. Then the price impact of one unit is $a_{m-1} - a_m$, which according to Theorem 1 equals $f_{m-2} - f_{m-1}$ (except for the case when $m = M$). The following result gives an asymptotic formula for the average price impact when $c > 1$.

Proposition 4. *In the context of Proposition 2, define $I_m = -\frac{\mathrm{d}f_{m-1}}{\mathrm{d}m}$ the price impact of a one-unit market order in state m. Then when competition $c > 1$,*

[31] One should not expect this conclusion to hold empirically, because it is not robust: the formula was derived in the one-sided model. In the two-sided model, the minimum bid-ask spread has a much more complicated dependence on the model parameters, and there is no reason why the dependence of the minimum bid-ask spread on competition c should even have the same sign.

the mean price impact $\bar{I}$ *and the standard deviation* $\sigma(I)$ *can be approximated by* ($\varepsilon = \frac{r}{\lambda}$)*:*

$$\bar{I} \approx \varepsilon \ln(1/\varepsilon) \frac{c+1}{\ln(c)}, \quad \sigma(I) \approx \sqrt{\varepsilon(A-B)} \sqrt{\frac{c+1}{c-1}}. \tag{12}$$

The average price impact decreases in λ*, and decreases in* c *as long as* $c < 3.5$.

Proof. See the Appendix. ■

Notice that price impact and the average bid-ask spread depend on granularity ε and competition c in the same way. This should not be surprising because both price impact and the bid-ask spread are forms of market depth (in the sense of Kyle 1985). When the market is deep (fast arrivals λ, large competition c), price impact is small, because all the inter-limit-order spreads are small as well.

The next result analyzes M, the maximum possible number of (sell) limit orders in the book. This number is endogenous, and it depends on the granularity ε, competition c, and the order book bounds A and B. It is important to determine M because it is related to the average density of the limit order book: a higher M describes a dense book, while a lower M describes a rarefied book.

Proposition 5. *In the context of Proposition* 2*, the maximum possible number* M *of sellers in the book is given by*

$$M = \frac{\ln\left(\frac{A-B}{\varepsilon} \frac{(c-1)^2}{c+1}\right)}{\ln c} + s, \quad \text{with } s \in \left(-1, \frac{\ln(2)}{\ln(c)}\right). \tag{13}$$

Formula (13) *is only true if granularity* ε *is sufficiently small—e.g., if* $\varepsilon \leq (A - B) \frac{c-1}{c+1} \frac{\ln(c)}{\ln(4c)}$. *The number* M *is increasing in trading activity* λ *and decreasing in competition* c.

Proof. See the Appendix. ■

It is surprising that the maximum number of sellers in the order book actually *decreases* with competition c (holding total activity λ constant). This is because of how the limit orders are placed in the book. When $c > 1$ is relatively large, the spreads between the different limit orders are smaller, but limit orders become more rarefied as one gets further away from the ask price. On balance, the maximum number of orders M is actually smaller when competition c is larger.[32]

[32] One may think that by a similar argument, the average bid-ask spread $\bar{S}$ should also be larger when c is large. This would be true if one used the *arithmetic* average across all states. However, here one takes a weighted average with weights given by stationary probabilities, and these are proportional to c^m. Because competition c is larger than one, smaller spreads are more likely than higher spreads, therefore the average spread decreases in c when c is not too high.

One may wonder which empirical implications of the one-sided model are robust—i.e., carry through to the general two-sided model. We think that the asymptotic formulas for the mean and standard deviation of the bid-ask spread, for the price impact, and for the maximum number of orders in the book M are true, except that one may have to multiply the results by a constant. Such a statement seems difficult to prove formally.

Finally, we discuss how to estimate the model parameters r, A, B, λ, and c. The patience coefficient r is not observed, so one may assume that it is constant for at least a short period of time and derive it from the implications of the model. One can argue also that r should be constant across stocks, as it must depend only on agents' type. The order book bounds A and B can be estimated simply by looking at the limits of a (perhaps winsorized) limit order book. Trading activity $\lambda = \lambda_1 + \lambda_2$ is also observable, as the sum of arrival rates of market and limit orders. A problem arises if one attempts to estimate competition $c = \lambda_1/\lambda_2$ as the ratio of arrival rates of sell limit orders to buy market orders. This is because, for example, λ_1 is the arrival rate of patient sellers, not of sell limit orders.[33] Instead, the model suggests estimating c as the average competition *conditional* on the various observed spreads. This is the same as the arithmetic average of the ratio λ_1/λ_2 over all states of the book, which can be shown to equal $cM/(M+1)$ and therefore is approximately equal to the theoretical value c.

4. Multi-Unit Market Orders and Price Impact

This section studies in more detail the price impact of a transaction, and the shape of the limit order book. The attention is focused on two questions: (i) If one submits a buy market order for i units, how much does the price change instantaneously? (ii) How does the limit order book change subsequently? The answer to the first question is given by studying the instantaneous (or temporary) price impact function. The answer to the second comes from analyzing the subsequent (or permanent) price impact.

Similar to the previous section, it is assumed that there are only patient sellers and impatient buyers. As in the discussion following Theorem 1, if only one-unit market orders are allowed, no limit orders above the ask are necessarily fixed in equilibrium, and so the price impact function is not uniquely defined. Therefore, to fix the other limit orders above the ask, the sellers should expect that their orders might be executed at any time with positive probability. This means that multi-unit market orders must be allowed.

[33] It is true that patient sellers typically submit limit orders, but there is an exception when the order book is full, in which case incoming patient sellers submit market orders. So one cannot estimate $c = \lambda_1/\lambda_2$ as the total number of sell limit orders divided by the total number of buy market orders, because that can be shown to theoretically equal one. In practice, however, this ratio is not equal to one due to cancellations, and to the extent that the number of cancellations is proportional to λ_1, one may argue that in fact this ratio is a good proxy for competition c.

4.1 Description of the equilibrium

As in the previous section, assume that patient sellers still arrive with only one unit to sell. The difference is that impatient buyers can submit up to k-unit market orders, with $k > 1$. Define

$$\begin{cases} \mu = \text{ arrival rate of patient sellers;} \\ \mu_i = \text{ arrival rate of } i\text{-unit impatient buyers, } i = 1, \ldots, k. \end{cases}$$

Assume that $\mu_i > 0$ for all $i = 1, \ldots, k$. As in the previous section, we concentrate on the case when the market is resilient—i.e., when patient sellers arrive faster than the units demanded by the impatient buyers (this way the bid-ask spread tends to revert to small values). This is equivalent to $\mu > \Sigma_{i=1}^{k} i\,\mu_i$. The number of states is finite, so there exists a maximum number M of limit orders in the book. Moreover, the expected utility of sellers when the book is full is $f_M = B$, the lower bound. From the state $m = 1, \ldots, M-1$ the system can go to one of the following states: $m+1$, if a patient seller arrives; or $m-i$, $i = 1, \ldots, k$ if an impatient i-buyer arrives. When there are M sellers, the bottom seller has a mixed strategy: after the first arrival of a Poisson process with intensity ν he places a market order at the reservation value, the lower bound B.

By calculations similar to the ones before Definition 1, the recursive system takes the following form:

$$\begin{cases} f_0 = f_{-1} = \cdots = f_{1-k} = A, \\ f_m = \frac{\mu}{\mu + \sum_{i=1}^{k} \mu_i} f_{m+1} + \sum_{i=1}^{k} \frac{\mu_i}{\mu + \sum_{i=1}^{k} \mu_i} f_{m-i} - \frac{r}{\mu + \sum_{i=1}^{k} \mu_i}, \quad 1 \le m \le M-1, \\ f_M = \frac{\mu_1 + \nu}{\sum_{i=1}^{k} \mu_i + \nu} f_{M-1} + \sum_{i=2}^{k} \frac{\mu_i}{\sum_{i=1}^{k} \mu_i + \nu} f_{M-i} - \frac{r}{\sum_{i=1}^{k} \mu_i + \nu}. \\ f_M = B. \end{cases} \tag{14}$$

The description of the equilibrium is the following.

Theorem 2. *Consider the limit order book with only patient sellers and impatient buyers ($\lambda_{PB} = \lambda_{IS} = 0$), with upper bound A, lower bound B; patient sellers arrive at rate μ and have patience coefficient r; impatient buyers who submit i-unit market orders arrive at rate μ_i, $i = 1, 2, \ldots, k$. Then there exists a Markov perfect equilibrium of the game, with a maximum number M of sellers, such that in the state with $m \le M$ patient sellers, the sellers' expected utility f_m is given by*

$$f_m = C_0 + C_1 \alpha_1^m + C_2 \alpha_2^m + \cdots + C_k \alpha_k^m + \frac{r}{\mu - \sum_{i=1}^{k} i\,\mu_i}\, m. \tag{15}$$

The complex numbers $\alpha_0 = 1, \alpha_1, \ldots, \alpha_k$ are the roots of the polynomial $P(X) = \mu X^{k+1} - (\mu + \sum_{i=1}^{k} \mu_i) X^k + \sum_{i=1}^{k} \mu_i X^{k-i}$, and the constants

$C_0, \ldots, C_k$ are determined uniquely from the first and the last equations of the recursive system (14).

Denote by $i_0 = \min\{k, m\}$. Then the level $a^i(m)$ of the ith limit order counted from bottom up, in the state with m sellers, for $i = 1, \ldots, i_0$, is

$$a^i(m) = \frac{\mu_k f_{m-k} + \mu_{k-1} f_{m-k+1} + \cdots + \mu_i f_{m-i}}{\mu_k + \mu_{k-1} + \cdots + \mu_i}, \tag{16}$$

where by convention $f_0 = f_{-1} = \cdots = f_{1-k} = A$.

The strategy of each agent in state m is the following: If $m = 1$, then place a limit order at $a^1(1) = A$. If $m = 2, \ldots, M-1$, look at the bottom k levels (or at all m levels if $m < k$), which are $a^1(m), \ldots, a^{i_0}(m)$; if any of them is not occupied, occupy it; anything above $a^{i_0}(m)$ does not matter. If $m = M$, the strategy is the same as for $m = 2, \ldots, M-1$, except for the bottom seller at $a^1(M)$, who exits (by placing a market order at B) after the first arrival in a Poisson process with intensity ν. If $m > M$, then immediately place a market order at B.

This equilibrium is unique in the class of rigid equilibria.

Proof. Very similar to the proof of Theorem 1. ■

One can now show that when i-unit market orders arrive at rates that are very small, the equilibrium of Theorem 2 converges to the canonical equilibrium from Definition 2, where each patient seller arriving in state m places an order at a_m, and the other sellers keep their orders unchanged.

Proposition 6. *The canonical equilibrium from Definition* 2 *is a limiting case of the equilibrium in Theorem* 2 *when all the market order arrival ratios μ_{i+1}/μ_i are very small.*

Proof. First one shows that $a^i(m) \approx f_{m-i}$: notice that in Equation (16) only the term with μ_i remains, since μ_i dominates all the other μ_j with $j > i$. The only thing left to prove is that f_m is approximately the same as the solution to the recursive system (1) for one-unit market orders. But this is easy to see, since μ_1 dominates all μ_i with $i > 1$, and therefore the recursive system (14) becomes approximately equal to Equation (1), where λ_1 is replaced by μ, and λ_2 is replaced by μ_1. ■

4.2 Price impact of transactions: Theory and empirical implications

Having described the equilibrium limit order book when there are multi-unit market orders, one can now define the temporary and permanent price impact functions. When an i-unit market order is submitted in a limit order book with m sellers, the first effect is that the market order clears the lowest i offers in the book. Therefore, the new ask price immediately becomes the $(i+1)$th lowest offer in the book. This leads to a temporary, or instantaneous, price impact. But

the agents are fully strategic, so they instantly regroup to adjust to the new state with $m - i$ sellers. The ask price therefore changes to a new level, which leads to a permanent, or subsequent, price impact.

The first main result of this section is that the temporary price impact is larger than the permanent price impact, which leads to price overshooting. The second result is that the price impact function essentially depends only on the relative arrival rates of market orders of different sizes.

Definition 3. *The* temporary (*or* instantaneous) *price impact function in a limit order book with m sellers maps each market order size i to the difference* $I_m(i) = a^{i+1}(m) - a^1(m)$ *between the* $(i+1)$*th offer* $a^{i+1}(m)$ *from the bottom and the first offer* $a^1(m)$ *from the bottom—i.e., the ask price.*

The permanent (*or* subsequent) *price impact function associates to each order size i the difference* $J_m(i) = a^1(m-i) - a^1(m)$ *between the ask price* $a^1(m-i)$ *in the state with* $m - i$ *sellers and the ask price* $a^1(m)$ *before the market order was submitted.*

One can compute the instantaneous price impact for the canonical equilibrium of the previous section, and see that it is convex in all cases.

Proposition 7. *The instantaneous price impact* $I_m(i)$ *for the canonical equilibrium of Definition* 2 *is a convex function of order size i.*

Proof. Use Proposition 11 to compute the second derivative of the expected utility f_m with respect to m. If $\lambda_1 \neq \lambda_2$, $\frac{\mathrm{d}^2 f_m}{\mathrm{d}m^2} = C(\frac{\lambda_2}{\lambda_1})^m \ln^2(\frac{\lambda_2}{\lambda_1}) > 0$; and if $\lambda_1 = \lambda_2$, $\frac{\mathrm{d}^2 f_m}{\mathrm{d}m^2} = \frac{2r}{\lambda_1+\lambda_2} > 0$. ■

To prove the main results of this section, one needs to assume that market order arrival rates are much larger for one unit than for two or more units.[34]

Assumption 1. For the rest of this section, assume that μ_1 is much larger than μ_i for $i > 1$. Equivalently, assume that

$$\mu_i = u\phi_i \quad \text{if } i > 1, \quad \text{with } u \text{ much smaller than } \mu_1, \text{ and } \phi_i \leq 1. \tag{17}$$

The numbers ϕ_i are called the *relative* arrival rates.

Proposition 8. *Under Assumption* 1, *prices overshoot—i.e., the temporary price impact* $I_m(i)$ *is larger than the permanent price impact* $J_m(i)$*. Also,* $I_m(i)$ *depends only on the relative arrival rates* ϕ_i.

Proof. One needs to show that $I_m(i) = a^{i+1}(m) - a^1(m) > a^1(m-i) - a^1(m) = J_m(i)$, which is equivalent to $a^{i+1}(m) > a^1(m-i)$. As in the proof of

[34] The price overshooting result probably holds under no restrictions, but I was not able to prove it in full generality.

Proposition 6, since μ_1 dominates μ_i for $i > 1$, one can show that $a^1(m-i) \approx f_{m-i-1}$. But Equation (16) implies that $a^{i+1}(m) = \frac{\mu_k f_{m-k} + \cdots + \mu_{i+1} f_{m-i-1}}{\mu_k + \cdots + \mu_{i+1}}$, which is a weighted average of $f_{m-k}, f_{m-k+1}, \ldots, f_{m-i-1}$, and therefore larger than f_{m-i-1}, since f_j is decreasing in j. It follows that $a^{i+1}(m) > a^1(m-i)$, hence that $I_m(i) > J_m(i)$.

To prove that $I_m(i)$ depends only on relative arrival rates, consider the formula $I_m(i) = a^{i+1}(m) - a^1(m) \approx \frac{\phi_k f_{m-k} + \cdots + \phi_{i+1} f_{m-i-1}}{\phi_k + \cdots + \phi_{i+1}} - f_{m-1}$. ■

The second result in Proposition 8 shows that limit order traders care mainly about relative arrival rates. If some seller is i levels above the ask, then either a market order has size less than i, in which case its arrival rate is not important, or it has size larger than i, in which case competition takes place only by considering conditional probabilities—i.e., only the relative arrival rates.

The first result is related to the empirical fact that the temporary price impact is larger than the permanent price impact (see Dejong, Nijman, and Roell 1996). The intuition in this model is that before an i-unit market comes, the sellers who expect their limit orders to be executed do not know the exact size, so in order to take advantage of larger orders, they stay higher in the book. Once the size becomes known, the sellers regroup lower in the book because of competition.[35] The next result derives a few empirical implications about price overshooting.

Proposition 9. *Consider a one-sided limit order book, with impatient buyers arriving at the rate $\mu_i = u\phi_i$ (as in Assumption 1), and patient sellers arriving at the rate $\mu > \sum_{i=1}^{k} i\mu_i$. Suppose that the relative arrival rates ϕ_i can be expressed as $\phi_i = \beta^i$, where $\beta \in (0,1)$. Define the competition parameter $c = \mu/\mu_1 > 1$, and assume that it is close to 1, in the sense that $\ln(c) \approx c - 1$. Suppose that the maximum size of market orders k is large, and the market order size i is much smaller than the number of sellers m. Then the relative price overshooting corresponding to order size i when there are m patient sellers approximately equals*

$$\frac{I_m(i) - J_m(i)}{I_m(i)} \approx \frac{\beta\left(c - 1 + \frac{1}{i}\right)}{1 - \beta c + \beta\left(c - 1 + \frac{1}{i}\right)}. \tag{18}$$

Relative price overshooting increases with β and competition c, decreases with order size i, and does not depend on number of sellers m.

Proof. See the Appendix. ■

The intuition is the following: Prices overshoot more when large market orders are relatively more likely (β is higher) because the patient sellers then

[35] Another interpretation of price overshooting is based on information. See Section 6 for a discussion about this alternative explanation.

have more incentive to stay higher in the book, which leads to a large reversal after the order size is revealed. Price overshooting increases with competition c because the reversal is due to competition. Prices overshoot less when the market order size i is higher: both price impacts $J_m(i)$ and $I_m(i)$ increase with i, but the temporary price impact $I_m(i)$ increases at a faster rate, due to the likelihood of larger orders. Finally, price overshooting does not depend on the number m of sellers, because it is defined as a relative measure. The absolute price overshooting $I_m(i) - J_m(i)$ can be seen in the proof of Proposition 9 to decrease with m: more sellers who compete in the limit order book reduce the overall size of price overshooting.

4.3 Price impact of transactions: Numeric results

In this section, we analyze numerically the shape of the limit order book, or, equivalently, the properties of the price impact function. What one usually calls the shape of the limit order book is a plot, which, on the horizontal axis, has the discrete grid of price levels above the ask, and on the vertical axis the depth existing at that level (i.e., the total size of sell limit orders at that price). Since the tick size is zero in this paper, by convention the depth at a certain discrete value is defined by collecting all the limit orders around that value.

Then what the literature calls a "hump-shaped" limit order book in the present model translates to the fact that limit orders cluster at some point above the ask.[36] The intuition is that patient sellers correctly anticipate that large multi-unit market orders are likely, and want to take advantage of them by submitting sell limit orders at higher prices. This translates by having a larger depth around the price where they cluster, thus creating a hump in the graph.

To calculate the price impact function in a concrete example, consider the following values for the numerical application: the upper bound of the limit order book is $A = 1$; the patience coefficient of the sellers is $r = 0.001$; the maximum market order size is $k = 20$; the arrival rates of impatient buyers are $\mu_1 = 1$; $\mu_i = u\phi_i$, $i = 2, \ldots, k$, where $u = 10^{-5}$, and the relative arrival rates ϕ_i are given below; and the arrival rate of the patient sellers is $\mu = \sum_{i=1}^{k} i\mu_i$.[37] Also, instead of defining the lower bound B, one can also define the first offer below A: $f_1 = A - 0.04$. For the relative arrival rates ϕ_i, we choose three types of decreasing functions: (i) $\phi_i = \frac{1}{i(i+1)}$ (quadratic); (ii) $\phi_i = \frac{1}{10^{i-1}}$ (exponential); and (iii) $\phi_i = \frac{1}{i(i+1)}$ if $i = 1, \ldots, 10$ and $\phi_i = \frac{1}{10^{i-9}\times 11}$ if $i = 10, \ldots, 20$ (mixed: first quadratic and then exponential).

Assume that the limit order book has at least $k = 20$ sellers—e.g., suppose that there are $m = 30$ sellers. The graphs are only displayed up to $k = 20$,

[36] See Biais, Hillion, and Spatt (1995) or Bouchaud, Mezard, and Potters (2002).

[37] Technically, in order to apply the results of this section, μ should be strictly larger than $\sum_{i=1}^{k} i\,\mu_i$. But it turns out that the analysis is the same in the case of equality.

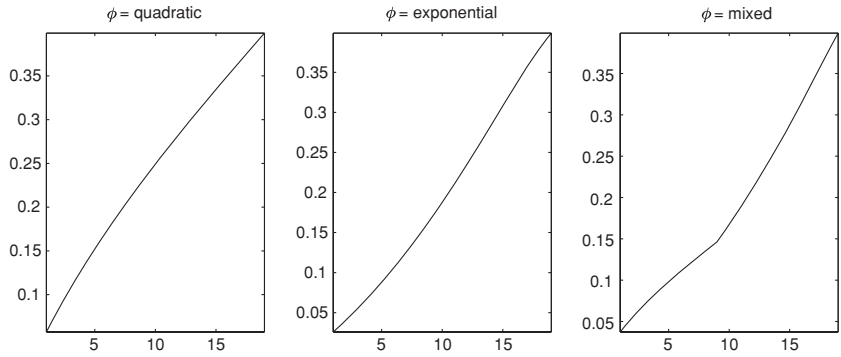

Figure 1
The instantaneous (temporary) price impact function $I_m(i)$ plotted against order size i when the relative market order arrival rates ϕ_i are: (i) $\phi_i = \frac{1}{i(i+1)}$ (quadratic); (ii) $\phi_i = \frac{1}{10^i - 1}$ (exponential); and (iii) $\phi_i = \frac{1}{i(i+1)}$ if $i = 2, \ldots, 10$ and $\phi_i = \frac{1}{10^{i-9} \times 11}$ if $i = 10, \ldots, 20$ (mixed: first quadratic and then exponential). $I_m(i)$ is the difference between the level of the ith limit sell order above the ask, and the ask price. The parameters are $A = 1, r = 0.001$, $f_1 = A - 0.04, k = 20$, and the arrival rates are $\mu_1 = 1; \mu_i = u\phi(i), i = 2, \ldots, k$, where $u = 10^{-5}$; $\mu = \sum_{i=1}^{k} i\mu_i$. The limit order book is in state $m = 30$ (the number of limit orders in the book). The graphs are drawn only up to $k = 20$, the maximum number of units that market buy orders can trade with positive probability. The shape of the equilibrium price impact function is: (i) concave, (ii) convex, and (iii) first concave then convex.

since market orders of size $i > k$ appear with zero probability, and so it does not matter what sellers do above the kth level from the ask.

The results confirm our intuition, as can be seen in Figure 1. If large market orders are more likely, as in the quadratic case (i), then the price impact function $I_m(i)$ is concave, which indicates that patient sellers indeed place their limit orders at higher levels above the ask. If large market orders are unlikely, as in the exponential case (iii), the price impact function is convex, indicating clustering of limit orders at the ask. For the mixed case, the price impact function is concave for $i \leq 10$ and convex for $i \geq 10$, which indicates clustering at a point above the ask.[38] This reflects the existing differences in the empirical literature, which has not yet said its final word whether the price impact is concave, linear, or convex, and in what range.[39]

5. Equilibrium: The General Case

Consider the general case, when all types of buyers and sellers arrive with positive probability. For simplicity, it is assumed that all the arrival rates are

[38] The exact predictions of the model are slightly different: as soon as the order reaches a certain size (equal to the existing number of limit orders in the order book), the price impact becomes flat. This is because it was assumed that as soon as prices reach A and B, an infinite number of agents from outside the model are willing to step in to supply liquidity at those prices.

[39] Huberman and Stanzl (2004) argue that a nonlinear price impact function leads to arbitrage. Empirical studies typically indicate a concave price function—e.g., Hasbrouck (1991, 1992); Keim and Madhavan (1996); and Knez and Ready (1996). If one eliminates the influence of large or block trades, the price impact function is almost linear—e.g., Breen, Hodrick, and Korajczyk (2002); and Sadka (2003).

equal: $\lambda = \lambda_{PB} = \lambda_{PS} = \lambda_{IB} = \lambda_{IS} > 0$. To get some intuition for the equilibrium, consider a setup similar to that of the one-sided case, but suppose that after the first patient seller (which has a limit sell order at A) a first patient buyer arrives in the market. Then the buyer behaves as a monopolist toward the potential incoming impatient sellers, and places a limit buy order at B. In this situation, if the reservation value of the seller is larger than the reservation value of the buyer, they will not be tempted to make offers to one another, and would rather wait to trade with future impatient agents. It follows that patient buyers and sellers behave very much as in the one-sided case, where new patient agents just keep placing spread-improving limit orders until it is better to trade immediately rather than wait.

Therefore, patient agents form two queues, a descending one with sell limit orders starting from A, and an ascending one with buy limit orders starting from B. Suppose now that the limit order book has m patient sellers and n patient buyers. Then, as in the one-sided case, all the m patient sellers have the same expected utility, say $f_{m,n}$, and and all the n buyers have the same expected utility, say $g_{m,n}$. Notice also that if traders prefer to wait, it must be that $f_{m,n} > g_{m,n}$.

When there is no more room in the limit order book—i.e., when the book is full—the traders on both sides play a game of attrition, in which they lose utility by waiting. As before, a *rigid* equilibrium is a competitive stationary Markov equilibrium in which only the bottom seller and the top buyer have mixed strategies. In that case, without loss of generality, the bottom seller places a limit order at some lower level h, and the top buyer immediately accepts the offer by placing a market order. Such a limit order is called *fleeting*. In the state where a fleeting order is placed, all traders, buyers and sellers, have the same expected utility $h = f_{m,n} = g_{m,n}$.

Unlike in the previous sections, in the general case even after restricting attention to rigid equilibria, one still does not get uniqueness. But one can still obtain *asymptotic* uniqueness: as the granularity parameter $\varepsilon = r/\lambda$ becomes very small, the equilibrium tends in the limit to the unique solution of a system of partial differential equations (see Theorem 4). This means that even though there might be multiple rigid equilibria, they are all very close to each other.

5.1 Description of the equilibrium

In the state with m sellers and n buyers, denote by $a_{m,n}$ the ask price, by $b_{m,n}$ the bid price, by $f_{m,n}$ the expected utility of the sellers, and by $g_{m,n}$ (minus) the expected utility of the buyers. As in the one-sided case, one can show that in a competitive stationary Markov equilibrium, the number of states (m, n) is finite. Then one defines the *state region* Ω as the collection of all states (m, n) where, in equilibrium, agents wait in expectation for some positive time. Also, one defines the *boundary* γ of Ω as the set of states in which at least some agent has a mixed strategy (and exits after Poisson(ν) for some arrival rate $\nu \geq 0$).

This is the set of states in which the limit order book becomes full, and it plays the same role as the maximum number M of limit orders in the one-sided case.

To be more precise, start with the parameters A, B, r, and λ, and define the granularity parameter $\varepsilon = r/\lambda$. Consider now some set Ω of pairs of integers in the positive quadrant, and let γ be its boundary. For each boundary point $(m, n) \in \gamma$ consider also a number $\nu_{m,n} \geq 0$. Since Ω should be the state space for an equilibrium, it must satisfy some extra properties, summarized in the following assumption.[40]

Assumption 2. *If (m, n) belongs to Ω, then $(m-1, n)$ and $(m, n-1)$ are also in Ω, as long as the coordinates are nonnegative. Moreover, on each 45° line (i.e., on each line parallel to the main diagonal) in the first quadrant that intersects Ω, there exists a unique point in γ.*

As in the one-sided case, it is a good idea to find a recursive structure for the expected utility functions f and g. From state $(m, n) \in \Omega$ the system can go to the following neighboring states: $(m-1, n)$, if an impatient buyer arrives, or if a patient seller cancels the limit order and submits a market order at B (when $n = 0$); $(m+1, n)$, if a patient seller arrives and submits a limit order; $(m, n-1)$, if an impatient seller arrives, or if a patient buyer cancels the limit order and places a market order at A (when $m = 0$); $(m, n+1)$, if a patient buyer arrives and submits a limit order; $(m-1, n-1)$, if after a positive expected time a seller places a quick (fleeting) limit order and a buyer immediately accepts.

Using arguments similar to those of Section 3, one gets a recursive system of equations in $f_{m,n}$ and $g_{m,n}$. The exact equations are collected in Definition 4 in the Appendix, together with the corresponding equations for the ask price $a_{m,n}$ and bid price $b_{m,n}$. The main result of this section is that, given a solution of the recursive system, there exists an equilibrium of the game.

Theorem 3. *Consider the limit order book with patient and impatient sellers and buyers, upper bound A, lower bound B. For simplicity, assume that all traders arrive at the same rate λ, and that patient traders have patience coefficient r. Consider a pair (Ω, ν), with Ω satisfying Assumption 2, and $\nu \geq 0$. Suppose that $f_{m,n}$, $g_{m,n}$ are a solution of the recursive system associated to the pair (Ω, ν) as in Definition 4 in the Appendix, and $a_{m,n}$, $b_{m,n}$ satisfy the formulas given in the same definition. Then there exists a competitive stationary Markov equilibrium of the game, which is also rigid, so that the state region is Ω, and the mixed strategies when the order book is full are given by ν.*

[40] Ideally, one would like to prove that any rigid equilibrium leads to a state region Ω, satisfying Assumption 2. This would be true if one could prove the following result: In any rigid equilibrium, the arrival of a new seller makes the sellers worse off and the buyers better off; moreover, the sellers are worse off by more than the buyers are better off.

To describe the equilibrium strategies, let m be the number of sellers and n the number of buyers in a state (m, n) not necessarily in Ω.[41] *If (m, n) is in Ω, but not in the boundary γ, the bottom seller places a limit sell order at $a_{m,n}$, and the top buyer places a limit buy order at $b_{m,n}$. If (m, n) is in γ, then the strategy is the same as the one above, except that after Poisson $(\nu_{m,n})$, the bottom seller changes the limit order from $a_{m,n}$ to $f_{m,n} = g_{m,n}$, and the top buyer immediately accepts via a market order; the top buyer would not accept any higher limit sell order. If (m, n) is not in Ω, and $m, n > 0$, then the bottom seller places a limit order at $f_{m,n} = g_{m,n}$ and the top buyer immediately accepts it via a market order. If (m, n) is not in Ω, and $n = 0$, then the bottom seller places a market order at B and exits the game.*

Proof. Similar to the proof of Theorem 1. ■

The theorem only guarantees the existence of an equilibrium. Uniqueness is a more delicate matter, and in the strict sense is not true, as can be seen numerically.[42] However, the solution is *asymptotically* unique: Theorem 4 below shows that, when granularity $\varepsilon = r/\lambda$ is small, all different equilibrium solutions $f_{m,n}$, $g_{m,n}$ are close to each other, since they all approach the *unique* solution of a system of partial differential equations.

Theorem 4. *In the context of Theorem* 3*, consider an equilibrium where patient sellers' expected utility is $f_{m,n}$ and patient buyers' expected utility is $g_{m,n}$. Denote the granularity $\varepsilon = \frac{r}{\lambda}$, and $\delta = \sqrt{\varepsilon}$. Define the functions f and g at the discrete values $(x, y) = (m\delta, n\delta)$ by $f(x, y) = f_{m,n}$, $g(x, y) = g_{m,n}$. Then when ε is small, f and g approach the solution of the following system of partial differential equations with free boundary γ:*

$$\begin{cases} \triangle f = 1, \\ f(0, y) = A, \\ \dfrac{\delta f}{\delta y}(x, 0) = 0, \\ \dfrac{\delta f}{\delta x} + \dfrac{\delta f}{\delta y} = 0 \text{ at } \gamma; \end{cases} \qquad \begin{cases} \triangle g = -1, \\ g(x, 0) = B, \\ \dfrac{\delta g}{\delta x}(0, y) = 0, \\ \dfrac{\delta g}{\delta x} + \dfrac{\delta g}{\delta y} = 0 \text{ at } \gamma; \end{cases} \tag{19}$$

where the free boundary γ is determined by the condition

$$f = g \text{ at } \gamma. \tag{20}$$

[41] As in the one-sided case, one needs to describe only the behavior of the bottom seller. If the bottom seller does not follow this strategy, then a seller above would immediately replace the bottom seller. Also, by symmetry, the strategy of the top buyer is similar.

[42] The reason for the multiplicity of equilibria is given by the complementarity between buyers and sellers in the boundary states γ. In general, one more seller makes the other sellers worse off by strictly more than the buyers are better off. But in the boundary states, this can happen with equality, which gives rise to the possibility of one extra state, with one more seller and one less buyer.

The problem is found numerically to have a unique solution, which is symmetric in x and y.[43] *The curve γ is convex and passes through the points* (1.96, 0) *and* (0, 1.96).

Proof. See the Appendix. ■

5.2 Empirical implications: The comovement effect and fleeting orders

To get more intuition for the two-sided equilibrium, it is helpful to imagine the limit order book as the collection of the sell side and the buy side, where each side has the reservation value given by the best limit order on the other side. For example, the sell side can be thought of as a one-sided limit order book where the limits are A and $B = b_{m,n}$, the bid price.[44]

As will be shown below, an implication of the equilibrium in the general case is that a market sell order leads to a decrease not only in the bid price—this could be explained by a mechanical execution of limit orders from the bid side—but also in the ask price. Moreover, the decrease in the bid price is larger than the decrease in the ask price, which leads to a wider bid-ask spread. We call this the *comovement* effect. It works in both directions: bid and ask prices tend to move together, with the side cleared by the market order moving first and by the larger amount.

The intuition for the comovement effect is the following: a market sell order clears the top-limit buy orders, which makes the buyers immediately better off, because there is less competition among them. At the same time, the sellers are worse off, because their reservation value (the bid price) went down. So they readjust and move down the ask price as well. But the decrease in reservation value does not directly affect the sellers until the limit order book becomes full. This means that decrease in the ask price should be smaller than the decrease in the bid price.

This phenomenon was noticed empirically by Biais, Hillion, and Spatt (1995), in their analysis of the order flow in the Paris Bourse (now Euronext). They attribute it to information: part of the decrease in the bid could be mechanical, but the decrease in the ask must be due to the information arising from the downward shift in the expected fundamental value. The present model generates this empirical regularity in a way that does not involve asymmetric information and, moreover, provides a way to estimate its magnitude.[45]

To illustrate the comovement effect in the present model, one can use a numerical example, described in Table 1. The parameter values are $A = 1$,

[43] Each partial differential equation is a Poisson equation in a closed region, with mixed-derivative conditions at the boundary. The condition $f = g$ at the boundary determines the free boundary γ, where the limit book is full. Since the oblique derivative is never tangent to γ, the problem is well posed, and one can write an algorithm to solve it, using finite differences. See, e.g., Gladwell and Wait (1979).

[44] To make the intuition rigorous, one would need to allow B to be stochastic in the one-sided model, and the depth at B to be just one unit.

[45] Another non-asymmetric information model that generates a comovement effect is Parlour (1998, Proposition 6).

Table 1
Solution in the general case with both buyers and sellers, for A = 1, B = 0, ε = 0.09

1.000	0.965					
1.000	0.905	0.824				
1.000	0.828	0.726	•			
1.000	0.770	0.616	0.500	•		
1.000	0.726	0.526	0.384	0.274	0.176	
1.000	0.697	0.468	0.300	0.177	0.095	0.035
1.000	0.682	0.440	0.260	0.131	0.045	0.000

$\nu = [0.21, 3.97, 0.99, 34.34, 2.50, 0.30, 3.47, 0.30, 2.50, 34.34, 0.99, 3.97, 0.21]$.

The left bottom corner corresponds to state (0, 0). The number in position (m, n) represents the expected utility $f_{m,n}$ of the sellers in state (m, n). The vector ν collects the variables corresponding to the mixed strategies along the boundary γ, starting from (0, 6) down to (6, 0) along γ. The expected utility $g_{m,n}$ of the buyers is given by the formula $g_{m,n} = 1 - f_{n,m}$. The bullets in positions (3, 4) and (4, 3), which are not in Ω, indicate the departure of the shape of Ω from the triangular shape.

$B = 0, \varepsilon = \frac{r}{\lambda} = 0.09$. The left bottom corner corresponds to state (0, 0). The number in position (m, n) represents the sellers' expected utility $f_{m,n}$ in the state with m sellers and n buyers. The vector ν collects the Poisson rates corresponding to the mixed strategies along the boundary γ, starting from (0, 6) down to (6, 0) along γ. This is a symmetric equilibrium—i.e., the expected utility $g_{m,n}$ for the buyers is given by the formula $g_{m,n} = 1 - f_{n,m}$.

Suppose that the limit order book is in state (2, 3), with two patient sellers and three patient buyers. In this state, the bid is $b_{2,3} = g_{2,2} = 1 - f_{2,2} = 0.474$, and the ask is $a_{2,3} = f_{1,3} = 0.770$. Then a market sell order for one unit moves the bid to $b_{2,2} = g_{2,1} = 1 - f_{1,2} = 0.274$, and the bid to $a_{2,2} = f_{1,2} = 0.726$. Also, a market sell order for two units moves the bid to $b_{2,1} = g_{2,0} = 1 - f_{0,2} = 0.000$ and the ask to $a_{2,1} = f_{1,1} = 0.697$. So while the bid moves from 0.474 to 0.274 to 0.000, the ask moves from 0.770 to 0.726 to 0.697.

The next proposition gives an approximate magnitude for the comovement effect.

Proposition 10. *Suppose that the limit order book is in the state with $m + 1$ sellers and n buyers. Assume that patient buyers and sellers arrive at the same rate λ_1, and impatient buyers and sellers arrive at the same rate λ_2, so that $\lambda_1 > \lambda_2$ are sufficiently large. Denote the competition parameter by $c = \frac{\lambda_1}{\lambda_2} > 1$. Then, if a market sell order moves the bid price down by Δ, the ask price moves down approximately by $\Delta(1 - \frac{1}{c^m})$. Therefore, the bid-ask spread increases approximately by $\frac{\Delta}{c^m}$.*

Proof. As mentioned above, the sell side of the book can be thought of as a one-sided model with $B = b_{m+1,n}$, the bid price. So when the bid price B moves down Δ, one needs to estimate the fall in the ask price $a_{m+1,n}$. Since $a_{m+1,n} = f_{m,n}$, the ask price falls approximately by $\Delta \times \frac{\mathrm{d}f_m}{\mathrm{d}B}$, with the dependence of f on n being omitted. Now, Proposition 1 of Section 3 implies that when $c > 1$, the derivative of f_m with respect to B in the one-sided case

is $\frac{\mathrm{d}f_m}{\mathrm{d}B} = (1 - \frac{1}{c^m})/(1 - \frac{1}{c^M})$. When M is large, which is true if trading activity $\lambda = \lambda_1 + \lambda_2$ is large enough, one can approximate $\frac{\mathrm{d}f_m}{\mathrm{d}B} \approx 1 - \frac{1}{c^m}$. So if the bid price B moves down by Δ, the ask price $a_{m+1,n}$ goes down by approximately $\Delta(1 - \frac{1}{c^m})$. ■

Notice that the comovement effect is stronger when there is more competition in the book: either when competition c is higher (future competition) or when the number of patient sellers $(m + 1)$ in the book is larger (current competition).[46]

Another consequence of the general case is the existence of quick, or *fleeting*, orders: when the limit order book becomes full, a buyer or seller will place a limit order, and a limit trader on the other side will immediately accept it by canceling the limit order and placing a market order. This model makes the strong prediction that fleeting orders should appear only when the order book is full (i.e., when the bid-ask spread is at its minimum), and they always improve the spread—i.e., are always placed between the bid and the ask.

6. Waiting Costs versus Information

This section compares the present model, based on waiting costs, with the model of Glosten (1994), which is based on asymmetric information and perfect competition. In Glosten's model, informed traders are assumed to always submit market orders, and uninformed traders are assumed to submit competitive limit orders at a price equal to the expected fundamental value conditional on the limit order being hit by a market order.[47]

For example, suppose that the limit order book has m patient sellers. Then the level a^i of the ith limit order above the ask price is $a^i = \mathsf{E}\{v \mid \text{buy market order of at least } i \text{ units}\}$. If μ_i is the probability of an i-unit buy market order, and k is the largest possible market order size, then one can derive the formula $a^i = \frac{\mu_k v_k + \mu_{k-1} v_{k-1} + \cdots + \mu_i v_i}{\mu_k + \mu_{k-1} + \cdots + \mu_i}$, where $v_j = \mathsf{E}\{v \mid \text{buy market order of exactly } j \text{ units}\}$. Notice that this is the same as Equation (16) in the multi-unit one-sided model of Section 4, except that the expected utility f_{m-j} is replaced by v_j. In Glosten (1994) there exists perfect competition, so the conditional value v_j should not depend on the number of sellers m. By contrast, in this model the

[46] The comovement effect does not depend on the number of patient buyers n, but this is because one assumes that the move Δ in the bid price is given. If instead one started with a market sell order of size i, then n would also show up in the formula for Δ. More precisely, if the number of competing buyers n is larger, the price impact Δ is smaller. In fact, one can see that $\Delta = J_m(i)$, the permanent price impact corresponding to an i-unit market order. See the proof of Proposition 9 in Section 4.2 for the exact formulas for the one-sided limit order book.

[47] It is not clear why patient informed traders would not want to trade via limit orders. Indeed, Ellul et al. (2007) provide evidence that orders routed to the automatic execution system of NYSE (as opposed to those routed to the floor auction process) display extreme impatience, yet they appear to be the less information-sensitive ones. Roşu (2008) allows informed traders to choose between limit and market orders and finds that they often prefer to trade via limit orders.

expected value f_{m-j} does depend on m, so, in principle, one could tell these two models apart.

But one would like to find a stronger way to distinguish between the two models, which does not assume that the information model is based on perfect competition. Luckily, there is a way to do that, and it is perhaps the most direct way to test the waiting costs assumption. Consider a limit order book with m offers and n bids. Suppose that the patient traders arrive at the same rate λ_1 and the impatient traders arrive at the same rate $\lambda_2 < \lambda_1$. Denote by $\lambda = 2(\lambda_1 + \lambda_2)$ the total trading activity. Then, as in the discussion before Theorem 3, one can write sellers' expected utility as $f_{m,n} = \frac{1}{\lambda}(\lambda_2 f_{m-1,n} + \lambda_1 f_{m+1,n} + \lambda_2 f_{m,n-1} + \lambda_1 f_{m,n+1}) - \frac{r}{\lambda}$. Notice that this looks like a formula based on the expectation of each type of order, except for the term $\frac{r}{\lambda}$, which is the loss in expected utility due to waiting costs.

Perhaps the best way to test this implication is to look at the bid-ask spread. The ask price $a_{m,n}$ satisfies $a_{m,n} = f_{m-1,n}$, the expected utility of sellers in state $(m-1, n)$, where there is one less seller. But $f_{m-1,n}$ satisfies a recursive equation, $f_{m-1,n} = \frac{1}{\lambda}(\lambda_2 f_{m-2,n} + \lambda_1 f_{m,n} + \lambda_2 f_{m-1,n-1} + \lambda_1 f_{m-1,n+1}) - \frac{r}{\lambda}$, so one gets $a_{m,n} = \frac{1}{\lambda}(\lambda_2 a_{m-1,n} + \lambda_1 a_{m+1,n} + \lambda_2 a_{m,n-1} + \lambda_1 a_{m,n+1}) - \frac{r}{\lambda}$. Similarly, the bid price equals $b_{m,n} = \frac{1}{\lambda}(\lambda_2 b_{m-1,n} + \lambda_1 b_{m+1,n} + \lambda_2 b_{m,n-1} + \lambda_1 b_{m,n+1}) + \frac{r}{\lambda}$ (recall that $b_{m,n} = g_{m,n-1}$, which is minus the expected utility of the buyers). If one defines the bid-ask spread $S_{m,n} = a_{m,n} - b_{m,n}$ in the state with m sellers and n buyers, one gets the formula

$$\begin{aligned} S_{m,n} + \frac{2r}{\lambda} &= \frac{1}{\lambda}(\lambda_2 S_{m-1,n} + \lambda_1 S_{m+1,n} + \lambda_2 S_{m,n-1} + \lambda_1 S_{m,n+1}) \\ &= \mathsf{E}\{S \mid \text{next order}\}, \end{aligned} \tag{21}$$

where $\mathsf{E}\{S \mid \text{next order}\}$ is the expected spread after the next order arrives in the state (m, n). This implies that, conditional on the system being in a generic state with m sellers and n buyers, the bid-ask spread should increase in expectation once the next order arrives. The intuition is that in order to compensate the patient traders who wait in the order book, the bid-ask spread must increase on average each period.

This does not mean that the bid-ask spread on average is wide: since patient traders arrive at a rate λ_1 higher than the arrival rate λ_2 of impatient traders, it is more likely that the book will be in a state with a smaller bid-ask spread. This more than compensates for the increase in the bid-ask spread, so the order book is in fact resilient, and the bid-ask spread tends to revert to the minimum value.

Therefore, a clear way to test the present model is to check whether the expected increase in the bid-ask spread conditional on the arrival of the next order (which equals $2\varepsilon = \frac{2r}{\lambda}$) is positive. If this fails empirically, one can still argue that the patience coefficient r is unknown, and perhaps it is too small to be detected using statistics. But this is unlikely: according to Corollary 1, the

expected increase in the bid-ask spread has the same order of magnitude as the minimum bid-ask spread, which is observable.

7. Conclusions

This paper presents a tractable model of the dynamics of the limit order book. The driving force is not asymmetric information, but waiting costs and competition among liquidity providers. Even though it is a stylized model, it delivers a rich set of implications about the shape of the limit order book and its evolution in time. Some implications of the model provide an alternative interpretation of existing empirical literature, and other implications are new. For example, higher trading activity and higher competition among traders cause smaller spreads and lower price impact. If large market orders are likely enough, the limit order book displays a "hump shape"—i.e., limit orders cluster away from the bid-ask spread. Prices also overshoot after market orders—i.e., the temporary price impact is larger than the permanent one. There is also a comovement effect for the bid and ask prices. Investors use quick (fleeting) limit orders when the limit order book is full. Finally, as seen in Section 6, the model can be tested against an alternative information model.

Appendix: Proofs of Results

An important ingredient in proving Theorem 1 is to show that the recursive system from Definition 1 has a unique solution.

Proposition 11. *There exists a unique solution* (f_m, M, ν) *to the associated recursive system from Definition 1. It satisfies the formulas* $(m = 0, \ldots, M)$

$$f_m = A + C\left(\left(\frac{\lambda_2}{\lambda_1}\right)^m - 1\right) + \frac{r}{\lambda_1 - \lambda_2}\, m, \quad \text{if } \lambda_1 \neq \lambda_2, \tag{A1}$$

$$f_m = A - bm + \frac{r}{\lambda_1 + \lambda_2}\, m^2, \quad \text{if } \lambda_1 = \lambda_2. \tag{A2}$$

The real numbers $C > 0$ *and* $b > 0$ *are given by*

$$C = \frac{r}{\lambda_1 - \lambda_2}\, \frac{\frac{\lambda_1 + \nu}{\lambda_2 + \nu}}{\left(\frac{\lambda_2}{\lambda_1}\right)^{M-1} - \left(\frac{\lambda_2}{\lambda_1}\right)^{M}}, \tag{A3}$$

$$b = \frac{2r}{\lambda_1 + \lambda_2}\left(M - \frac{\nu - \lambda_1}{2(\nu + \lambda_1)}\right), \tag{A4}$$

where M *is the unique positive integer that for some* $\nu \geq 0$ *satisfies*

$$\frac{A - B}{\frac{r}{\lambda_1 - \lambda_2}} = \frac{\lambda_1 + \nu}{\lambda_2 + \nu}\, \frac{\left(\frac{\lambda_1}{\lambda_2}\right)^M - 1}{\left(\frac{\lambda_1}{\lambda_2}\right) - 1} - M, \quad \text{if } \lambda_1 \neq \lambda_2, \tag{A5}$$

$$M = \frac{\nu - \lambda_1}{2(\nu + \lambda_1)} + \sqrt{\left(\frac{\nu - \lambda_1}{2(\nu + \lambda_1)}\right)^2 + \frac{A - B}{\frac{r}{\lambda_1 + \lambda_2}}}, \quad \text{if } \lambda_1 = \lambda_2. \tag{A6}$$

Also, if f_m is extended for $m > M$ via the above formula, then f_m is strictly decreasing in m if $m < M$ and strictly increasing if $m > M$.

Proof. Let $\lambda = \lambda_1 + \lambda_2$, $a = \frac{\lambda_1}{\lambda_2}$, and $\varepsilon = \frac{r}{\lambda}$. Start with the case $c \neq 1$. One needs to solve the difference equation: $(\lambda_1 + \lambda_2) f_m + r = \lambda_1 f_{m+1} + \lambda_2 f_{m-1}$, or $f_m + \varepsilon = \frac{c}{c+1} f_{m+1} + \frac{1}{c+1} f_{m-1}$. Defining $g_m = f_m - f_{m-1}$, one must now solve the homogeneous equation $g_m = \frac{c}{c+1} g_{m+1} + \frac{1}{c+1} g_{m-1}$. The characteristic equation is $x = \frac{c}{c+1} x^2 + \frac{1}{c+1}$, which has two roots: $x_1 = 1$ and $x_2 = \frac{1}{c}$. Therefore, the general solution is $g_m = C_1 \cdot 1^m + C_2 \cdot x_2^m$, where C_1 and C_2 are some arbitrary real numbers. But $f_m = f_0 + g_1 + g_2 + \cdots + g_m = A + C_1 m + C_2 \frac{x_2^{m+1} - x_2}{x_2 - 1}$. Requiring now that f_m satisfies $f_m + \varepsilon = \frac{c}{c+1} f_{m+1} + \frac{1}{c+1} f_{m-1}$ for $m = 0$, one gets $C_1 = \frac{\varepsilon}{(c-1)/(c+1)} = \frac{r}{\lambda_1 - \lambda_2}$. Defining $C = C_2 \frac{x_2}{x_2 - 1}$, one finally obtains $f_m = A + C(x_2^m - 1) + \frac{r}{\lambda_1 - \lambda_2} m$. One also needs to impose that $f_M = f_{M-1} - \frac{r}{\lambda_2 + \nu}$, which implies $C = \frac{r}{\lambda_1 - \lambda_2} \frac{\lambda_1 + \nu}{\lambda_2 + \nu} \frac{1}{c^{-(M-1)} - c^{-M}}$.

The integer M is determined by the requirement that $f_M = B$, which leads to $\frac{A - B}{\frac{r}{\lambda_1 - \lambda_2}} = \frac{\lambda_1 + \nu}{\lambda_2 + \nu} \frac{c^M - 1}{c - 1} - M$. Define $S = \frac{A - B}{\frac{r}{\lambda_1 - \lambda_2}} > 0$. It is elementary to show that when $c > 1$ and $\nu \in [0, \infty]$, the right-hand side of the equation equals zero at $M = 1$ and is strictly increasing in $M > 1$. This means that for any $S > 0$, there is a unique $M_\nu > 1$ that solves the equation. When $\nu = \infty$, the equation is $S = \frac{c^{M_\infty} - 1}{c - 1} - M_\infty$. This can be rewritten as $S = \frac{\lambda_1}{\lambda_2} \frac{c^{M_\infty - 1} - 1}{c - 1} - (M_\infty - 1)$. But M_0 satisfies $S = \frac{\lambda_1}{\lambda_2} \frac{c^{M_0} - 1}{c - 1} - M_0$, so $M_0 = M_\infty - 1$. As M_ν is strictly decreasing in ν, it follows that there is a unique $\nu \in [0, \infty)$ such that M_ν is an integer. This fixes ν and completes the proof.

In the case when $c < 1$, the same proof works, except that the equation is now (multiplying the previous one by -1) $\frac{A - B}{\frac{r}{\lambda_2 - \lambda_1}} = M - \frac{\lambda_1 + \nu}{\lambda_2 + \nu} \frac{1 - c^M}{1 - c}$. The proof now goes along the same lines as before.

Now consider the case when $c = 1$. Then one needs to solve the difference equation: $(\lambda_1 + \lambda_2) f_m + r = \lambda_1 f_{m+1} + \lambda_2 f_{m-1}$, or $f_m + \varepsilon = \frac{1}{2} f_{m+1} + \frac{1}{2} f_{m-1}$. Defining $g_m = f_m - f_{m-1}$, one must now solve the homogeneous equation $g_m = \frac{1}{2} g_{m+1} + \frac{1}{2} g_{m-1}$. The characteristic equation is $x = \frac{1}{2} x^2 + \frac{1}{2}$, which the double root $x = 1$. Therefore the general solution is $g_m = C_1 + C_2 m$, where C_1 and C_2 are some arbitrary real numbers. But $f_m = f_0 + g_1 + g_2 + \cdots + g_m = A + C_1 m + C_2 \frac{m(m+1)}{2}$. Requiring now that f_m satisfies $f_m + \varepsilon = \frac{1}{2} f_{m+1} + \frac{1}{2} f_{m-1}$ for $m = 0$, one gets $C_2 = 2\varepsilon = \frac{2r}{\lambda_1 + \lambda_2}$. Defining $b = -(C_1 + C_2/2)$, one finally obtains $f_m = A - bm + \frac{r}{\lambda_1 + \lambda_2} m^2$. One also needs to impose that $f_M = f_{M-1} - \frac{r}{\lambda_2 + \nu}$, which implies $b = 2\varepsilon(M - \frac{1}{2} + u)$, where $u = \frac{\lambda_1}{\nu + \lambda_1} \in (0, 1]$.

The integer M is determined by the requirement that $f_M = B$, which leads to $M^2 - 2M(\frac{1}{2} - u) - \frac{A - B}{\varepsilon} = 0$. This has the unique positive solution $M_u = \frac{1}{2} - u + \sqrt{(\frac{1}{2} - u)^2 + \frac{A - B}{\varepsilon}}$. One also has $M_0 - M_1 = 1$, so there is a unique $u \in (0, 1]$ such that M_u is an integer. This leads to the desired formulas.

Finally, let us prove the last statement of the proposition in the case $c = 1$; the proof is similar for the other cases. The function f_m is quadratic, and it is first decreasing in m, then increasing in m. To determine where $f'(m)$ changes signs, solve $f'(m^*) = 0$, i.e., $b = \varepsilon m^*$. This gives $m^* = M - \frac{1}{2} + u$, which belongs to the interval $(M - \frac{1}{2}, M + \frac{1}{2}]$. This shows that f is strictly decreasing if $m < M$ and strictly increasing if $m > M$. ∎

Before proving Theorem 1, it is important to understand what happens in the various states of the equilibrium, when there is a fixed number of sellers in the limit order book.

Proposition 12. *Suppose that m sellers lose utility in a way proportional to expected waiting time with coefficient r. At random time T, which represents the first arrival in a Poisson process with intensity λ, an event happens and the game ends (this event can be the arrival of a new agent).*

Then, if all the sellers wait until T, assume that each gets a payoff of f^∞. Also, at each time there exists a buyer who posts a bid for h. Assume that if a seller accepts h until T, he gets h and all other sellers get f^-. Denote $f^0 = f^\infty - r/\lambda$. Then one has the following list of possible subgame perfect equilibria:

1. *If $h > \max\{f^0, f^-\}$, then every seller immediately accepts h (and only one randomly gets it).*
2. *If $h < \min\{f^0, f^-\}$, then no seller accepts h, and everybody waits until T.*
3. *If $h \in [f^-, f^0]$ and $f^- < f^0$, there are two SPE:*
   (a) *Each seller waits until T.*
   (b) *Each seller places a market order for h (if they believe the others will try to get h, they are all better off doing the same).*
4. *If $h \in [f^0, f^-]$ and $f^0 \leq f^-$, this is a typical game of attrition. It has two equilibria:*
   (a) *Some agent always accepts h, and the others never accept h.*
   (b) *All agents accept h according to some Poisson process with intensity ν (ν is such that each agent is indifferent between accepting h now and waiting for the other $m-1$ sellers to do that).*

Proof of Proposition 12. Cases 1 and 2 are obvious. In Case 3, if all agents could coordinate and wait until the end, they would all be better off (and get utility f^0, which is greater than both f^- and h). However, if some agent deviates and accepts h, then everyone else gets $f^- \leq h$, so they would be better off by rushing to accept h as well.

Case 4 is a typical game of attrition: nobody wants to wait until the end (f^0 is smaller than both h and f^-), but at the same time nobody really likes to drop from the race and accept h, because h is less than the utility f^- they would get if someone else dropped out. The fact that only equilibria of type 4a and 4b exist is standard. See, e.g., Fudenberg and Tirole (1991, Section 4.5.2). ■

In the context of Theorem 1, behavior of type 1 appears in states $m > M$; behavior of type 2 appears in states $m = 1, \ldots, M-1$; and behavior of type 4 appears in state $m = M$. Notice that the previous result does not assume anything about sellers placing limit orders. The next result is a simple extension of this game of attrition, where sellers are allowed to place limit orders. Clearly, the ask price is important now, because that might influence the payoff f^∞ at T. It is shown that one gets two more equilibria.

Corollary 2. *In the setup of Proposition* 12*, assume that the sellers place limit orders as in the context of Theorem* 1*. Also, the event that ends the game is the arrival of an impatient buyer, which immediately places a market order. If everybody waits until then, assume that the top sellers get utility f^∞, while the bottom seller gets the ask price. Define f^0 as in Proposition* 12*. Then, if $h \in [f^0, f^-]$ and $f^0 \leq f^-$, besides the equilibria in Proposition* 12*, there exist two more equilibria:*

4. (c) *The top sellers wait, and the bottom seller randomly accepts h with Poisson(ν), where ν is defined such that the top sellers' utility in equilibrium is h. The ask price is defined such that the bottom seller's utility is also h.*
   (d) *Similar to* 4c*, except that the bottom seller waits, and the top sellers randomly accept h with Poisson(ν).*

These equilibria describe how agents behave in the states where the limit order book is full—i.e., when $m = M$. Recall that a competitive stationary Markov equilibrium is called *rigid* if the behavior of agents in state $m = M$ is of type 4c.

Proof of Corollary 2. The new fact here is that the bottom seller can influence his payoff at T by changing the ask price. That makes the bottom seller different from the top sellers. Now, clearly

either all the top sellers randomize their strategy, or none of them does (because they mix between the same values). So there are four cases: one in which no seller randomizes (Case 4a), one in which all sellers randomize (Case 4b), one in which only the bottom seller randomizes (Case 4c), and one in which only the top sellers randomize (Case 4d). ■

Proof of Theorem 1. The proof of existence is straightforward: one needs to show that the strategies described by the theorem lead to a competitive stationary Markov equilibrium. This is done as indicated in the beginning of Section 3.2.

To prove uniqueness, it is enough to show that any rigid equilibrium must be of the form described in the theorem. For this, one first shows that in such an equilibrium all agents have the same utility function. According to Roşu (2006), strategies must satisfy Property A4 (they must have finitely many jumps), which implies that the state variables are left-continuous. In particular, it makes sense to talk about the value of the state variables (m, a_m) right before some time t.

Consider a restriction of the strategies to some time interval $(t, t+\delta)$, such that no strategy has a jump during that interval. This restriction can be made because of the Markov condition: history is reduced to the limit of outcomes at a single point. On this interval, all agents have the same utility: Otherwise, suppose that the bottom seller is worse off than a top seller. Then the bottom seller can bid just a little bit higher, and he will achieve a higher utility. Now, suppose that the bottom seller is better off than a top seller. Then the top seller can "undercut by a penny," so she would be strictly better off than before. Therefore, the top and bottom sellers have the same utility.

Notice that this utility does not depend on the state variable a_m (the ask price), since agents' decisions are forward looking. The only problem would be if some agent's placing an order at a_m would prevent others from placing their desired orders. But this does not happen in the present case, since all agents have the same utility, and therefore order positions are interchangeable. This shows that the utility of the sellers depends only on the state variable m. Denote this utility by f_m.

It is clear that there are only a finite number of states m in which agents wait for a positive expected time: agents lose utility proportionally to expected waiting time, and in equilibrium their expected utility has to be larger than the reservation value B. Define M to be the largest state in which agents wait at least for a positive expected time.

Next, it is shown that the utility of each agent in the largest state M has to be exactly $f_M = B$. The case $f_M < B$ is not possible, because then the agent would not wait in that state. Suppose that $f_M > B$. Then consider what happens in state $M+1$. As in Proposition 12, if one agent accepted $h = B$ and exited the game, the utility of the other agents would be $f^- = f_M > B$. Recall that f^0 is the utility of the sellers if everybody waits. One can have either $f^0 > B$ or $f^0 \leq B$. If $f^0 > B$, then h is lower than both f^0 and f^-, so we are in Case 2, when all sellers wait. But this is in contradiction with M being the largest state in which agents wait. If $f^0 \leq B$, this is Case 4 of the proposition. Since the equilibrium is rigid, agents wait in this state, which again is in contradiction with the definition of M. This shows that $f_M = B$.

In a similar way, one can prove by induction that $f_m = B$ for all states $m \geq M$. In state $m = M$, the system goes either to state $M-1$ or to $M+1$, so one calculates $f^0 = \frac{1}{2}(f_{M-1} + B) - \frac{\varepsilon}{2}$. The utility $f^- = f_{M-1} \geq B$. It is easy to see that $f^- = f_{M-1} \geq f^0$. The case $h = B < f^0$ cannot occur, because then all agents would wait until the end in state M and get utility $f_M = f^0 > B$, which is in contradiction with $f_M = B$. If $h = B \geq f^0$, only case 4c in Corollary 2 can occur, because the equilibrium is assumed to be rigid. This is indeed the behavior prescribed by the theorem. Notice that in state $m = M$ all agents have utility $f_M = B$. Moreover, if $f^- = B$, agents do not wait at all in state M, which is in contradiction with the definition of M. Since $f^- \geq B$, it must be that $f^- = f_{M-1} > B$.

Now focus on state $m = M-1$. As before, one can show that $f^- = f_{M-2} > f^0$. If $h = B \geq f^0$, the rigidity of the equilibrium implies that this is Case 4c of Corollary 2, so $f_{M-1} = B$, a contradiction with what was just proved before ($f^- = f_{M-1} > B$). Then it follows that $h = B < f^0$, and all agents wait until the end. This is exactly the behavior prescribed by the theorem in

the cases when $m < M$. Another formula that comes from the analysis of the state $m = M - 1$ is $h < f^0 < f^-$; i.e., $B < f_{M-1} < f_{M-2}$. By induction, one can extend the above reasoning to all states $m < M$. This completes the proof of uniqueness. ■

Proof of Proposition 2. One needs to prove the following equations, depending on the competition parameter c:

$$\bar{S} \approx \varepsilon \ln(1/\varepsilon) \frac{c(c+1)}{(c-1)\ln(c)}, \quad \sigma(S) \approx \sqrt{\varepsilon(A-B)} \left(\frac{(c+1)(c^3+c^2-c)}{(c-1)^3} \right)^{1/2}, \quad \text{if } c > 1, \tag{A7}$$

$$\bar{S} \approx \frac{A-B}{3}, \quad \sigma(S) \approx \frac{2(A-B)}{3}, \quad \text{if } c = 1, \tag{A8}$$

$$\bar{S} \approx A - B, \quad \sigma(S) \approx \varepsilon \frac{(c+1)\sqrt{1-3c+4c^2-c^4}}{(1-c)^2}, \quad \text{if } c < 1. \tag{A9}$$

For simplicity, assume that the trader at the ask does not have a mixed strategy—i.e., that $\nu = 0$. The one-sided market with different arrival rates is a Markov system with transition matrix

$$P = \begin{bmatrix} \frac{1}{c+1} & \frac{c}{c+1} & 0 & 0 & \cdots & 0 & 0 \\ \frac{1}{c+1} & 0 & \frac{c}{c+1} & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{c+1} & 0 & \frac{c}{c+1} & \cdots & 0 & 0 \\ 0 & 0 & \frac{1}{c+1} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & \frac{c}{c+1} \\ 0 & 0 & 0 & 0 & \cdots & \frac{1}{c+1} & \frac{c}{c+1} \end{bmatrix}. \tag{A10}$$

To calculate the distribution of the bid-ask spread, one needs to know the stationary probability that the system is in state m. Denote this by x_m. Consider the row vector X with entries x_m. From the theory of Markov matrices, one knows that $XP = X$. Solving for X, one gets $\frac{1}{c+1} x_{m+1} = \frac{c}{c+1} x_m$ for all m, hence $x_m = Dc^m$. The components x_m must sum to one, so $D = \frac{c-1}{c^{M+1}-1}$ when $c \neq 1$, and $D = \frac{1}{M+1}$ when $c = 1$.

To calculate the average bid-ask spread $a_m - B$, notice that the spread is $A - B$ in state $m = 0$, $f_{m-1} - B$ in state $m = 1, \ldots, M-1$, and $\frac{r}{\lambda_2}$ in state $m = M$. To compute then the mean and standard deviation of the spread, one needs to perform a tedious calculation that will be omitted. To give just one example, when $c > 1$, the average spread can be computed to be precisely $\overline{a_m - B} = \frac{r/\lambda\,(c+1)}{(c-1)^2(c^{M+1}-1)}((M-1)c^{M+3} - (M-2)c^{M+2} - 3c^{M+1} + c^2 + (M+1)c - M) \approx \frac{r/\lambda\,c(c+1)}{c-1} M$. But from Proposition 5, $M \approx \frac{\ln(\frac{1}{r/\lambda}) + \ln((A-B)\frac{(c-1)^2}{c+1})}{\ln(c)} \approx \frac{\ln(\frac{1}{r/\lambda})}{\ln(c)}$. Since $\varepsilon = \frac{r}{\lambda}$, one gets $\overline{a_m - B} \approx \frac{c(c+1)\varepsilon \ln(\frac{1}{\varepsilon})}{(c-1)\ln(c)}$. ■

Proof of Proposition 4. Similar to the proof of Proposition 2, one needs to consider the average price impact over states m that appear with probabilities proportional to c^m. One computes $I_m = f_{m-2} - f_{m-1} = -\frac{r}{\lambda_1 - \lambda_2} + C((\frac{\lambda_2}{\lambda_1})^{m-2} - (\frac{\lambda_2}{\lambda_1})^{m-1}) = \frac{r}{\lambda_1 - \lambda_2}(c^{M+2-m} - 1)$, and, after some tedious calculations, one gets the desired formulas. ■

Proof of Proposition 5. One needs to prove the following equations, depending on the competition parameter c:

$$M = \frac{\ln\left(\frac{A-B}{r/\lambda}\,\frac{(c-1)^2}{c+1}\right)}{\ln c} + s, \quad \text{with} \quad s \in \left(-1, \frac{\ln(2)}{\ln(c)}\right), \qquad \text{if} c > 1, \tag{A11}$$

$$M = \sqrt{\frac{A-B}{r\lambda}} + s, \quad \text{with} \quad s \in (-1, 1), \qquad \text{if c=1}, \tag{A12}$$

$$M = \frac{A-B}{r/\lambda}\,\frac{1-c}{1+c} + s, \quad \text{with} \quad s \in \left(1, \frac{2-c}{1-c}\right), \qquad \text{if} c < 1. \tag{A13}$$

Consider first the case $c = \frac{\lambda_1}{\lambda_2} > 1$. For simplicity, assume that $\nu = \infty$ (all M_ν are within an interval of length one). According to Proposition 11, $M = M_\infty > 1$ solves the equation $\frac{c^M-1}{c-1} - M = S = \frac{A-B}{r/\lambda_1-\lambda_2} = \frac{A-B}{r/\lambda}\,\frac{c-1}{c+1} > 0$. This is the same as $c^M = 1 + (c-1)M + (c-1)S$, or $M = \log_c(1 + (c-1)M + (c-1)S)$. Denote by $S' = (c-1)S$. Define the sequence x_i recursively by $x_0 = 0$ and $x_{i+1} = \log_c(1 + (c-1)x_i + S')$. One can see that $x_i > \log_c(1 + S')$. By induction, it is shown that $x_i \le \log_c(1 + 2S')$. Suppose that one showed this inequality up to i. Then $x_{i+1} \le \log_c(1 + (c-1)\log_c(1 + 2S') + S')$, so if one shows that $(c-1)\log_c(1 + 2S') \le S'$, the proof is complete. This is equivalent to $\log_c(1 + 2(c-1)S) \le S$, or $2S \le \frac{c^S-1}{c-1}$. This is true if S is large enough: one can check that the desired inequality is true if $S \ge 1 + \log_c(4)$. This shows that $\log_c(1 + (c-1)S) \le M_\infty \le \log_c(1 + 2(c-1)S) \le \log_c(1 + (c-1)S) + \log_c(2)$. But all $M = M_\nu$ are between $M_\infty - 1$ and M_∞. So $\log_c(1 + (c-1)S) - 1 < M < \log_c(1 + (c-1)S) + \log_c(2)$ for $S \ge 1 + \log_c(4)$, which proves the desired result.

For the case $c = 1$, one computes $M_{1/2} = \sqrt{\frac{A-B}{r/\lambda}}$. But $M \in (M_{1/2} - 1, M_{1/2} + 1)$, which is what had to be proved.

When $c < 1$, M_∞ solves $M - \frac{1-c^M}{1-c} = T = \frac{A-B}{r/\lambda_2-\lambda_1} > 0$. So $M_\infty = T + \frac{1-c^{M_\infty}}{1-c}$, which together with $M_\infty > 1$ implies $M_\infty \in (T + 1, T + \frac{1}{1-c})$. But any M belongs to $(M_\infty, M_0 = M_\infty + 1]$, so $M \in (T + 1, T + \frac{2-c}{1-c})$.

One can show that when competition $c > 1$, the average number of limit traders $\overline{m}$ is approximately equal to M (i.e., $(M - \overline{m})/M$ converges to zero when M is large). For this, one uses the methods from the proof of Proposition 2. The state m appears with probability proportional to c^m, so $\overline{m} = \frac{1\times0+c\times1+c^2\times2+\cdots+c^M\times M}{1+c+c^2+\cdots+c^M} = \frac{c(Mc^{M+1}-(M+1)c^M+1)}{(c-1)(c^{M+1}-1)} = M - o(M)$, where $o(M)/M \to 0$ when $M \to \infty$.

Since $S = \frac{A-B}{r/\lambda}\,\frac{c-1}{c+1}$ increases in λ, it is clear that M increases in λ. To show that M decreases in c for λ sufficiently large, consider the asymptotic formula $M \approx \frac{\ln(S(c-1)^2/c+1)}{\ln(c)} = h(c)$. The derivative $\frac{dh}{dc} = \frac{c^2+3c/c^2-1-h(c)}{c\ln(c)}$, which is negative if $h(c)$ is sufficiently large—i.e., if S (or λ) is sufficiently large. ■

Proof of Proposition 9. In the course of proving Proposition 8, we saw that the permanent and temporary price impacts are given by $J_m(i) = a^1(m-i) - a^1(m) = f_{m-i-1} - f_{m-1}$, $I_m(i) = a^{i+1}(m) - a^1(m) = \frac{\phi_k(f_{m-k}-f_{m-1})+\cdots+\phi_{i+1}(f_{m-i-1}-f_{m-1})}{\phi_k+\cdots+\phi_{i+1}}$. As in the proof of Proposition 6, since μ_1 is much larger than μ_i for $i > 1$ (by Assumption 1), one can approximate f_m by the formula in Equation (2), where λ_1 is replaced by μ, and λ_2 is replaced by μ_1. Define the competition parameter $c = \mu/\mu_1$. Now, using Equation (2), one derives $f_{m-i-j} - f_{m-1} = C(c^{i+j-m} - c^{1-m}) + \frac{r}{\mu-\mu_1}(1 - i - j)$, $j \ge 1$. The term $\frac{r}{\mu-\mu_1}(1 - i - j)$ is much smaller than $C(c^{i+j-m} - c^{1-m})$, since m does not appear anymore in the former (it is assumed that i is much smaller than m), and so it is omitted. One gets the approximate formulas: $I_m(i) \approx Cc^{1-m}((\frac{(\beta c)^i-(\beta c)^k}{1-\beta c})/(\frac{\beta^i-\beta^k}{1-\beta}) - 1)$, and $J_m(i) \approx Cc^{1-m}(c^i - 1)$. Since $c > 1$ is assumed to be close to 1 and $\beta \in (0, 1)$, it follows that $\beta c < 1$. Now assume that k is large, which implies that the terms involving k in the formula for $I_m(i)$ drop. One gets $J_m(i)/I_m(i) = (c^i - 1)/(c^i\,\frac{1-\beta}{1-\beta c} - 1)$.

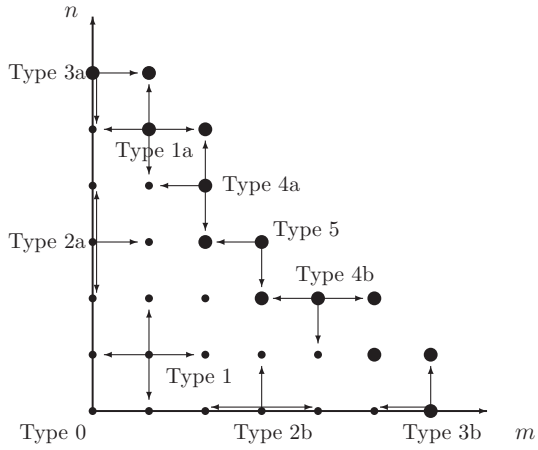

Figure 2
Types of points in the state region Ω

Since c is close to 1, $c^i \approx 1 + i(c-1)$, so one gets $J_m(i)/I_m(i) = (1-\beta c)/(1-\beta+\frac{\beta}{i})$. This implies $(I_m(i) - J_m(i))/I_m(i) = \beta(c - 1 + \frac{1}{i})/(1-\beta+\frac{\beta}{i})$. ■

In the two-sided limit order book, for any state region Ω that satisfies Assumption 2 in Section 5.1, one can define various types of points in Ω as in Figure 2: types 0, 1, 1a, 2a, 2b, 3a, 3b, 4a, 4b, 5.[48]

Definition 4. *Consider a region Ω in the positive quadrant that satisfies Assumption 2—i.e., if (m,n) is in Ω, then $(m-1,n)$ and $(m,n-1)$ are also in Ω, as long as they belong to the positive quadrant. For each boundary point $(m,n) \in \gamma$, consider a number $\nu_{m,n} \geq 0$. Let ν be the collection of all $\nu_{m,n}$. Then define the recursive system associated to (Ω, ν) by considering for each state $(m,n) \in \Omega$ the following set of equations: If (m,n) is of Type 0, $f_{0,0} = A$, $g_{0,0} = B$. If (m,n) is of Type 1, $4f_{m,n} + \varepsilon = f_{m-1,n} + f_{m+1,n} + f_{m,n-1} + f_{m,n+1}$, $4g_{m,n} - \varepsilon = g_{m-1,n} + g_{m+1,n} + g_{m,n-1} + g_{m,n+1}$. If (m,n) is of Type 1a, $(4+\frac{\nu_{m,n}}{\lambda})f_{m,n} + \varepsilon = f_{m-1,n} + f_{m+1,n} + f_{m,n-1} + f_{m,n+1} + \frac{\nu_{m,n}}{\lambda} f_{m-1,n-1}$, $(4+\frac{\nu_{m,n}}{\lambda})g_{m,n} - \varepsilon = g_{m-1,n} + g_{m+1,n} + g_{m,n-1} + g_{m,n+1} + \frac{\nu_{m,n}}{\lambda} g_{m-1,n-1}$, $f_{m,n} = g_{m,n}$. If (m,n) is of Type 2a, $f_{0,n} = A$, $3g_{0,n} - \varepsilon = g_{0,n-1} + g_{0,n+1} + g_{1,n}$; and similarly for Type 2b. If (m,n) is of Type 3a, $f_{0,n} = A$, $(2+\frac{\nu_{0,n}}{\lambda})g_{0,n} - \varepsilon = (1+\frac{\nu_{0,n}}{\lambda})g_{0,n-1} + g_{1,n}$, $f_{0,n} = g_{0,n}$; and similarly for Type 3b. If (m,n) is of Type 4a, $(4+\frac{\nu_{m,n}}{\lambda})f_{m,n} + \varepsilon = f_{m-1,n} + 2f_{m,n-1} + f_{m,n+1} + \frac{\nu_{m,n}}{\lambda} f_{m-1,n-1}$, $(4+\frac{\nu_{m,n}}{\lambda})g_{m,n} - \varepsilon = g_{m-1,n} + 2g_{m,n-1} + g_{m,n+1} + \frac{\nu_{m,n}}{\lambda} g_{m-1,n-1}$, $f_{m,n} = g_{m,n}$; and similarly for Type 4b. If (m,n) is of Type 5, $(4+\frac{\nu_{m,n}}{\lambda})f_{m,n} + \varepsilon = 2f_{m-1,n} + 2f_{m,n-1} + \frac{\nu_{m,n}}{\lambda} f_{m-1,n-1}$, $(4+\frac{\nu_{m,n}}{\lambda})g_{m,n} - \varepsilon = 2g_{m-1,n} + 2g_{m,n-1} + \frac{\nu_{m,n}}{\lambda} g_{m-1,n-1}$, $f_{m,n} = g_{m,n}$.*

If (m,n) is not in Ω, and $m, n > 0$, consider the unique point (m',n') in γ that lies on the 45° line that passes through (m,n). Then define $f_{m,n}$ and $g_{m,n}$ by the corresponding values at (m',n'). If the 45° line does not intersect γ, but it intersects one of the coordinate axes, simply define $f_{m,n} = g_{m,n}$ to be either A or B depending on whether it is the x-axis or the y-axis. Finally, if (m,n) is not in Ω, and $n = 0$, define $f_{m,n} = g_{m,n} = B$; and, similarly, when $m = 0$.

[48] There are two more types of boundary points in Ω. For example, in Figure 2, assume that Ω contains two extra points, of coordinates $(7,0)$ and $(8,0)$. These points lead to two different types of recursive equations, but it turns out that they cannot exist in equilibrium. Therefore, to simplify the discussion, these types of points will be ignored.

Also, given a solution of the recursive system, define a set of numbers $a_{m,n}$ and $b_{m,n}$ by the following formulas: If (m, n) is of Type 1, $a_{m,n} = f_{m-1,n}$, $b_{m,n} = g_{m,n-1}$. If (m, n) is of Type 2a, $a_{0,n} = A$, $b_{0,n} = g_{0,n-1}$; and similarly for Type 2b. If (m, n) is of Type 5, then for some $\frac{\nu_{m,n}}{\lambda} \geq 0$, $a_{m,n} = f_{m-1,n} + \frac{\nu_{m,n}}{\lambda}(f_{m-1,n-1} - f_{m,n})$, $b_{m,n} = g_{m,n-1} + \frac{\nu_{m,n}}{\lambda}(g_{m-1,n-1} - g_{m,n})$. The formulas for the other types of boundary points are similar.

Proof of Theorem 4. First show that $\triangle f = 1$. Start with equation $4f_{m,n} + \varepsilon = f_{m-1,n} + f_{m+1,n} + f_{m,n-1} + f_{m,n+1}$, and divide throughout by $\varepsilon = \delta^2$. Then one gets $\frac{f_{m-1,n}-2f_{m,n}+f_{m+1,n}}{\delta^2} + \frac{f_{m,n-1}-2f_{m,n}+f_{m,n+1}}{\delta^2} = 1$. This is the finite difference approximation of $(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2})(m\delta, n\delta) = 1$, which is the desired PDE: $\triangle f(x, y) = 1$.

Now, equation $3f_{m,0} + \varepsilon = f_{m-1,0} + f_{m+1,0} + f_{m,1}$ becomes, after division by δ, $\frac{f_{m-1,0}-2f_{m,0}+f_{m+1,0}}{\delta^2} \cdot \delta + \frac{f_{m,1}-f_{m,0}}{\delta} = \delta$. After passing to the limit when δ goes to zero, one gets $\frac{\partial f}{\partial y}(x, 0) = 0$.

If one picks a point on γ of type 1, one has $(4 + \frac{\nu_{m,n}}{\lambda})f_{m,n} + \varepsilon = 2f_{m-1,n} + 2f_{m,n-1} + \frac{\nu_{m,n}}{\lambda} f_{m-1,n-1}$, which after division by δ becomes $2\frac{f_{m,n}-f_{m-1,n}}{\delta} + 2\frac{f_{m,n}-f_{m-1,n}}{\delta} + \frac{\nu_{m,n}}{\lambda}\frac{f_{m,n}-f_{m-1,n-1}}{\delta} = -\delta$. After passing to the limit when δ goes to zero, one gets $\frac{\partial f}{\partial x}(x, y) + \frac{\partial f}{\partial y}(x, y) = 0$.

For a point on γ of type 2, one has $(4 + \frac{\nu_{m,n}}{\lambda})f_{m,n} + \varepsilon = f_{m-1,n} + 2f_{m,n-1} + f_{m,n+1} + \frac{\nu_{m,n}}{\lambda} f_{m-1,n-1}$, which becomes $\frac{f_{m,n}-f_{m-1,n}}{\delta} + \frac{f_{m,n}-f_{m-1,n}}{\delta} + \frac{f_{m,n-1}-2f_{m,n}+f_{m,n+1}}{\delta^2} \cdot \delta + \frac{\nu_{m,n}}{\lambda}\frac{f_{m,n}-f_{m-1,n-1}}{\delta} = -\delta$. In the limit one gets the same condition $\frac{\partial f}{\partial x}(x, y) + \frac{\partial f}{\partial y}(x, y) = 0$.

Finally, the condition $f = g$ on γ is obvious.

References

Bergin, J., and W. B. MacLeod. 1993. Continuous Time Repeated Games. *International Economic Review* 34:21–37.

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* 50:1655–89.

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

Bloomfield, R., M. O'Hara, and G. Saar. 2005. The "Make or Take" Decision in an Electronic Market: Evidence on the Evolution of Liquidity. *Journal of Financial Economics* 75:165–99.

Bouchaud, J.-P., M. Mézard, and M. Potters. 2002. Statistical Properties of the Stock Order Books: Empirical Results and Models. *Quantitative Finance* 2:251–56.

Breen, W. J., L. S. Hodrick, and R. A. Korajczyk. 2002. Predicting Equity Liquidity. *Management Science* 48:470–83.

Chakravarty, S., and C. W. Holden. 1995. An Integrated Model of Market and Limit Orders. *Journal of Financial Intermediation* 4:213–41.

Christie, W. G., and P. H. Schultz. 1994. Why Do NASDAQ Market Makers Avoid Odd-Eight Quotes? *Journal of Finance* 49:1813–40.

Cohen, K. J., S. F. Maier, R. A. Schwartz, and D. K. Whitcomb. 1981. Transaction Costs, Order Placement Strategy, and Existence of the Bid-Ask Spread. *Journal of Political Economy* 89:287–305.

de Jong, F., T. Nijman, and A. Röell. 1996. Price Effects of Trading and Components of the Bid-Ask Spread on the Paris Bourse. *Journal of Empirical Finance* 3:193–213.

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

Easley, D., and M. O'Hara. 1987. Price, Trade Size, and Information in Securities Markets. *Journal of Financial Economics* 19:69–90.

Ellul, A., C. W. Holden, P. Jain, and R. Jennings. 2007. Order Dynamics: Recent Evidence from the NYSE. Working Paper, Indiana University, January 2007.

Farmer, J. D., P. Patelli, and I. I. Zovko. 2005. The Predictive Power of Zero Intelligence in Financial Markets. *Proceedings of the National Academy of Sciences, USA* 102(6):2254–59.

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

Foucault, T., O. Kadan, and E. Kandel. 2005. Limit Order Book as a Market for Liquidity. *Review of Financial Studies* 18:1171–217.

Fudenberg, D., and J. Tirole. 1991. *Game Theory*. Cambridge, MA: MIT Press.

Gladwell, I., and R. Wait. 1979. *A Survey of Numerical Methods for Partial Differential Equations*. Oxford: Clarendon Press.

Glosten, L. 1994. Is the Electronic Open Limit Order Book Inevitable? *Journal of Finance* 49:1127–61.

Glosten, L., and P. Milgrom. 1985. Bid, Ask, and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders. *Journal of Financial Economics* 13:71–100.

Goettler, R. L., C. A. Parlour, and U. Rajan. 2005. Equilibrium in a Dynamic Limit Order Market. *Journal of Finance* 60:2149–92.

Handa, P., and R. A. Schwartz. 1996. Limit Order Trading. *Journal of Finance* 51:1835–61.

Harris, L., and J. Hasbrouck. 1996. Market vs. Limit Orders: The SuperDOT Evidence on Order Submission Strategy. *Journal of Financial and Quantitative Analysis* 31:213–31.

Hasbrouck, J. 1991. Measuring the Information Content of Stock Trades. *Journal of Finance* 46:179–207.

Hasbrouck, J., and G. Sofianos. 1993. The Trades of Market Makers: An Empirical Analysis of NYSE Specialists. *Journal of Finance* 48:1565–93.

Hausman, J. A., A. W. Lo, and A. C. MacKinlay. 1992. An Ordered Probit Analysis of Transaction Stock Prices. *Journal of Financial Economics* 31:319–30.

Hollifield, B., R. A. Miller, and P. Sandås. 2004. Empirical Analysis of Limit Order Markets. *Review of Economics and Statistics* 71:1027–63.

Hollifield, B., R. A. Miller, P. Sandås, and J. Slive. 2006. Estimating the Gains from Trade in Limit-Order Markets. *Journal of Finance* 61:2753–804.

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

Huberman, G., and W. Stanzl. 2004. Price Manipulation and Quasi-arbitrage. *Econometrica* 72:1247–75.

Jain, P. 2003. Institutional Design and Liquidity on Stock Exchanges around the World. Working Paper, University of Memphis, December 2003.

Jones, C. M., G. Kaul, and M. L. Lipson. 1994. Transactions, Volume, and Volatility. *Review of Financial Studies* 4:631–51.

Keim, D. B., and A. Madhavan. 1996. The Upstairs Market for Large-Block Transactions: Analysis and Measurement of Price Effects. *Review of Financial Studies* 9:1–36.

Knez, P. J., and M. J. Ready. 1996. Estimating the Profits from Trading Strategies. *Review of Financial Studies* 9:1121–63.

Kyle, A. 1985. Continuous Auctions and Insider Trading. *Econometrica* 53:1315–36.

Linnainmaa, J., and I. Roşu. 2008. Time Series Determinants of Liquidity in a Limit Order Market. Working Paper, University of Chicago.

Lo, A. W., A. C. MacKinlay, and J. Zhang. 2002. Econometric Models of Limit-Order Executions. *Journal of Financial Economics* 65:31–71.

O'Hara, M. 1995. *Market Microstructure Theory*. Oxford: Blackwell Publishers.

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

Rock, K. 1996. The Specialist's Order Book and Price Anomalies. Working Paper.

Roşu, I. 2006. Multi-stage Game Theory in Continous Time. Working Paper.

Roşu, I. 2008. Liquidity and Information in Limit Order Markets. Working Paper.

Sadka, R. 2003. Momentum, Liquidity Risk, and Limits to Arbitrage. Job Market Paper, Northwestern University.

Sandås, P. 2001. Adverse Selection and Competitive Market Making: Empirical Evidence from a Limit Order Market. *Review of Financial Studies* 14:703–34.

Seppi, D. J. 1997. Liquidity Provision with Limit Orders and a Strategic Specialist. *Review of Financial Studies* 10:103–50.

Simon, L. K., and M. B. Stinchcombe. 1989. Extensive Form Games in Continuous Time: Pure Strategies. *Econometrica* 57:1171–214.

Wyart, M., J.-P. Bouchaud, J. Kockelkoren, M. Potters, and M. Vettorazzo. 2008. Relation between Bid-Ask Spread, Impact and Volatility in Order-Driven Markets. *Quantitative Finance* 2:251–56.