A DYNAMIC MODEL OF THE LIMIT ORDER BOOK

By Ioanid Rosu[1]

I propose a continuous-time model of price formation in a market where trading is conducted according to a limit-order book. Strategic liquidity traders arrive randomly in the market and dynamically choose between limit and market orders, trading off execution price with waiting costs. I prove the existence of a Markov equilibrium in which the bid and ask prices depend only on the numbers of buy and sell orders in the book, and which can be characterized in closed-form in several cases of interest. My model generates empirically verified implications for the shape of the limit-order book and the dynamics of prices and trades. In particular, I show that buy and sell orders can cluster away from the bid-ask spread, thus generating a hump-shaped limit-order book. Also, following a market buy order, both the ask and bid prices increase, with the ask increasing more than the bid—hence the spread widens.

Keywords: Liquidity, price impact, limit order market, waiting costs, continuous time game theory, game of attrition.

1. Introduction

This article presents a model of price formation in a market where agents trade via a limit order book.[2] In such a market, there is no market maker or specialist who provides bid and ask quotes. Instead, any sellers (buyers) can place offers (bids) in the

Date: April 25, 2005.

[1]The author thanks Rob Battalio, Shane Corwin, Drew Fudenberg, Xavier Gabaix, Larry Glosten, Burton Hollifield, Sergei Izmalkov, Leonid Kogan, Jon Lewellen, Andrew Lo, David McAdams, David Musto, Stew Myers, Jun Pan, Anna Pavlova, Christine Parlour, Duane Seppi, Chester Spatt, Richard Stanton, Dimitri Vayanos, and Jiang Wang for very helpful comments and suggestions. He is also grateful to participants at the NBER market microstructure meeting, May 2004, and to seminar audiences at MIT, UC Berkeley, Notre Dame, U Toronto, Northwestern, Carnegie Mellon, U Michigan, U Penn, and U Chicago.

[2]The limit 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*, 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*.

limit order book and wait until the orders get executed; or, alternatively, they can trade immediately by placing a market order against the existing bids (offers).

The study of price formation when market makers do not exist or only have a limited role is very important in understanding modern financial markets. Nowadays, many financial markets around the world are order-driven, with a limit order book at the center of the trading process.[3] A satisfactory model of order-driven markets should therefore explain how market prices arise from the interaction of a large number of anonymous traders, who arrive at the market at random times, can choose whether to trade immediately or to wait, and can behave strategically by changing their orders at any time.

In this paper, I propose a model of an order-driven market which reflects the features mentioned above. The model is tractable and produces sharp implications about (i) the shape of the limit order book at any point in time, and (ii) the evolution in time of the book, and in particular of the bid and ask prices. Some of these implications are in line with known empirical facts about the limit order book and its dynamics. The determinants of price formation in this model are: the speed of agents' arrival to the market, their waiting costs, and the ratio of the numbers of patient to impatient traders.[4]

The model represents a departure from classical market microstructure. In that literature prices change because the suppliers of liquidity have to protect themselves from traders with superior information (Glosten and Milgrom (1985), Kyle (1985)). In particular, the bid-ask spread (and the price impact of a transaction) should be higher when there is more asymmetric information in the market. By contrast, in the present model prices change because the arrival of new agents modifies the balance between the

[3]About half of the world's stock exchanges are organized as order-driven markets, with no designated market makers (see Jain (2002)). Examples include Euronext, Hong Kong, Tokyo, Toronto, and various ECN (Island, Instinet). There are also hybrid exchanges (NYSE, Nasdaq, London), where market makers exist but have to compete with other traders, who supply liquidity by limit orders. In these markets, the number of transactions which involve a market maker is usually small (see Hasbrouck and Sofianos (1993).

[4]These variables have been showed empirically to explain a large percentage of the variation in the average bid-ask spread (Farmer, Patelli, and Zovko (2003)). Some evidence that reasons other than information may better explain prices can also be found for example in Huang and Stoll (1997), who estimate that on average approximately 90% of the bid-ask spread is due to non-informational frictions ("order-processing costs").

various suppliers of liquidity. In particular, the bid-ask spread should be higher when agents arrive more slowly to the market. In that case, the whole limit order book is more rarefied, because the traders must wait longer. Put differently, a market is traditionally considered liquid if the amount of asymmetric information in the market is small, while in my framework a market is liquid if it is fast.

In some sense, this paper reverts from the economics of information to the classical framework of supply and demand, but with a few changes that make that framework more realistic: there is no Walrasian auctioneer, agents are allowed to be strategic, and supply and demand are revealed dynamically, via limit and market orders. From this perspective, my model can in principle be applied to other types of markets, as long as the buy or sell prices are determined by strategic suppliers of liquidity.

Specifically, I consider a continuous-time, infinite-horizon economy where there is only one asset with no dividends. Buyers and sellers arrive at the market randomly. They either buy or sell one unit of the asset, after which they exit the model. I assume 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 over when to trade, and whether to place a market or limit order. After a limit order is placed, it can be canceled and 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.

In equilibrium, impatient agents submit market orders, while patient agents submit limit orders except for the states where the limit order book is "full." In those states 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 or the sellers. In states where the book is not full, new limit orders are always placed inside the bid-ask spread.[5] The point where the limit order book is full coincides with the time when

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

the bid-ask spread is at the minimum. That there exists a non-zero 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 are simply not allowed to place limit orders). This gives intuition for the general case by describing the sell side of the limit order book.

The case when there are no impatient agents can also be solved in closed-form. It turns out that in that case buyers and sellers cannot coexist in the limit order book: there are either many sellers in the book and the buyers place market orders, or vice versa. This shows that the existence of impatient agents (or rather of agents of different patience) is important in order for a limit order book to function properly.

In order to discuss price impact and determine the distribution of limit orders in the book, I allow for the possibility that multi-unit market orders arrive with positive probability. I show that if the such orders arrive with probabilities which do not decrease too fast with order size, then the book exhibits a hump shape, i.e., the limit orders cluster away from the bid and the ask.[6]

Solving the general case is more difficult, but it has the interesting empirical implication that after a market buy order not only the ask price increases, but also the bid price.[7] The fact that the bid also increases is documented for example in Biais et al. (1995), and has been taken as evidence of the presence of asymmetric information. I show that even in the absence of information such dynamics may occur.

The limit order book has been analyzed in a variety of ways. A precursor of this literature is the "gravitational pull" model of Cohen, Maier, Schwartz and Whitcomb (1981), where traders base choose between limit and market orders based on their expectations about the evolution of an exogenous price process. The information models, which consider market makers interacting with informed agents, are all static: see Glosten (1994), Chakravarty and Holden (1995), Rock (1996), Seppi (1997) and Parlour and

[6]This effect was observed among others by Biais, Hillion and Spatt (1995), and Bouchaud, Mezard and Potters (2002).

[7]The ask increases mechanically, because an offer is cleared from the book. The fact that the bid increases as well reflects the buyers' realization that their reservation value (the ask price) has moved further away, so they also adjust their bids.

Seppi (2001). Moreover, traders are restricted to placing limit orders, so they do not have a choice to submit market orders. Dynamic models, without market makers, are studied by Parlour (1998), Foucault (1999), Foucault, Kadan and Kandel (2004), Goettler, Parlour and Rajan (2004). However, these models are typically not very tractable, and do not allow for strategic cancellation of limit orders.

Although the present paper was developed independently from this literature, it turns out that it is closely related to the work of Foucault, Kadan, and Kandel (2004). Their setup is similar, with the exception that in their model traders are not allowed to cancel limit orders, and prices can only take a set of discrete values. In order to solve the model, they only focus on the bid-ask spread, and make two restrictive assumptions regarding the trading process: (i) a buyer must always arrive after a seller, and vice versa; and (ii) new orders have to be improving the existing limit orders by at least a tick.

A related literature is on liquidity and search costs, e.g., see Duffie, Garleanu, and Pedersen (2004), Vayanos and Wang (2003). In these models, buyers and sellers have to search for counter-parties to trade. My contention is that on organized exchanges search costs mostly become one-dimensional and can be better thought of as waiting costs.[8]

The paper is organized as follows. Section 2 describes the model. Section 3 solves for the equilibrium in a particular case which represents the sell side of the book: there are only patient sellers and impatient buyers, and the two types arrive at equal rates. Section 4 solves the similar problem when patient sellers arrive faster than the impatient buyers. Section 5 analyzes the case of multi-unit market orders and applies the results to analyze price impact and the shape of the limit order book. Section 6 describes the equilibrium in the general case with all types of sellers and buyers. Section 7 solves for the equilibrium in the homogeneous case when there are only patient agents, and Section 8 concludes.

[8] In some cases, search costs are important, e.g., in the upstairs market for large (block) trades, or when stocks are traded on more than one exchange.

2. The Model

2.1. *The Market*

In this section I present the assumptions of the model. For a brief discussion about these assumptions, see Section 2.2. Consider a market in an asset which 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 $A > B$ where the prices lie at all times. More specifically, there is an infinite supply when price is A, provided 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.[9] Limit orders can be canceled for no cost at any time. 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.[10]

Agents. The market is populated by traders who arrive randomly at the market. The arrival process is assumed exogenous, and will be described in more detail below. Once traders arrive, they choose strategically between market and limit orders. They are liquidity traders, in the sense that they want to trade the asset for reasons exogenous to the model. 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 exit the model forever.

[9]To justify this assumption, it is best to think of a market buy/sell order as a (marketable) limit order with limit 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 remains as limit orders, which can be freely canceled.

[10]For example, on Nasdaq the Level II system displays the best bids and offers from market makers and ECNs, and is publicly available to registered traders. On NYSE the limit order book is public (with a 5-second delay), but orders from the trading floor and stop-loss orders are only visible to the specialist. Also, on Euronext or the Toronto stock exchange, traders can place hidden limit orders, out of which only a small fraction is publicly visible.

Traders are risk-neutral, so their instantaneous utility function (felicity) is linear in price. By convention, felicity is equal to price for sellers, and minus the price for buyers. Traders discount the future in a way proportional to the expected waiting time.[11] If τ is the random execution time and P_τ is the price obtained at τ, the expected utility of a seller is

$$f_t = \mathsf{E}_t\{P_\tau - r(\tau - t)\}.$$

(The expectation operator takes as given the strategies of all the players, including Nature. See the description of strategies below.) Similarly, the expected utility of a buyer is $-g_t = \mathsf{E}_t\{-P_\tau - r(\tau - t)\}$, where I introduce the notation

$$g_t = \mathsf{E}_t\{P_\tau + r(\tau - t)\}.$$

I call f_t the value function, or utility, of the seller at t; and similary g_t the value function, or utility, of the buyer, although in fact g_t equals minus the expected utility of a buyer.

The discount coefficient r is constant, and can take two values: if it is low, the corresponding traders are called *patient*, otherwise they are *impatient*. Agents' types are determined from the beginning and cannot change.

For simplicity, I assume that the impatient agents always submit market orders. One can remove this assumption without much difficulty. In Remark 3 of Appendix A, I show that the equilibrium obtained would be essentially the same. Therefore, from now on I denote by r only the time discount coefficient of the patient agents.

Arrivals. The four types of traders (patient buyers, patient sellers, impatient buyers, and impatient sellers) arrive at the market according to independent Poisson processes with constant arrival intensity rates

$$\lambda_{PB}, \lambda_{PS}, \lambda_{IB}, \lambda_{IS}.$$

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

[11]This model also works with exponential time discounting, but the resulting formulas are slightly more complicated.

same parameter λ. The mean time until the next arrival is then $1/\lambda$. The interval until the next arrival is called a *period*.

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

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.

But there are costs to setting the game in continuous time. One is that there is no universally accepted standard of continuous time game theory. The main difficulty comes from the fact that given a time t, there is no last time before t, and no first time after t. Because of that, strategies may not map uniquely into outcomes (see the discussion in Appendix B.) There are two methods of dealing with this problem: one introduced by Simon and Stinchcombe (1989), who define a continuous-time game as a limit of discrete-time games; the other, pioneered by Bergin and MacLeod (1993) in their analysis of repeated games in continuous time, focuses on strategies with infinitesimal inertia. I find the latter framework to be the more appropriate for this paper, and I extend it in three directions.

First, suppose that in our model an impatient buyer submits a market order at time t. What is the ask price at which that order is executed? Since there is no last time before t, one needs to have a well defined notion of the outcome of the game immediately before t. This means that one should use strategies that do not behave too wildly. The technical concept, inspired from Simon and Stinchcombe (1989), is of a strategy with a uniformly bounded number of jumps (to be defined in Appendix B).

Also, when a market order arrives at time t, an existing limit trader exits the model, and the next stage of the game will be played with fewer traders. But at which time will this next game be played? No $t + \varepsilon > t$ is satisfactory, because it would imply agents waiting for a positive time, during which they lose utility. The best solution is to "stop the clock," so that the next game is also played at time t. The clock is restarted only when in the stage game no agent submits a market order. Allowing for clock stopping in continuous time game theory requires some care, and it is done in Appendix B.

The last extension I consider in order to define stochastic multi-stage games in continuous time is that of mixed strategies. Unlike discrete time, in continuous time mixing can be done both over actions, and over time (choosing the time of an action). 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. In the rest of the paper, I only consider equilibria where mixing is done over time.

The types of equilibrium used are subgame perfect equilibrium, and Markov perfect equilibrium (see Fudenberg and Tirole, ch. 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 behavior in the rest of the game does not change). All these notions are discussed in more detail in Appendix B.

Also, for the purposes of this paper, I also introduce 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). In the language of Corollary 13, in case 4 only equilibria of type c) occur.

Finally, in this paper all information, together with agents' strategies and beliefs are common knowledge.

2.2. *Discussion*

An important assumption is that the arrival of traders is exogenous. One reason for it is methodological: in many cases economic analysis takes certain variables such as income, or prices as exogenous, even though they clearly are not. Another reason is that even if individual traders know when they themselves arrive at the market, they typically do not know what the others will do. Attempts to explain the large observed trading volume using measurable information so far have not produced satisfactory results. Therefore, it might be a good strategy to remain agnostic as to why new traders arrive at the market, and instead focus on their interaction once they do arrive.

One may question the assumption that prices lie within a fixed exogenous range $[B, A]$, and that A and B are known by everybody with certainty. Where do A and B come from? One can think about them as summarizing information about the asset: $(A+B)/2$ would represent the average value of an asset, while $(B-A)$ would represent differences of opinion among traders. A more realistic assumption would be to make A and B stochastic, perhaps as prices coming from valuations of informed traders. Thus they can change every time new public information arrives. Ideally, a model with private valuations would explain not only A and B, but also where the the order flow comes from (not all of it, some order flow should still be exogenous, due to liquidity needs). Such generalizations are beyond the scope of this paper, and are left for future research.

Another strong assumption is that agents only trade one unit of the asset, after which exit the model. One may argue that this is not realistic. For example, some agents may decide to stay in the market and buy and sell securities, thus in effect becoming market makers.[12] Similarly, speculators may try to hoard liquidity and make monopoly profits. Also, what about the existence of large trades? A partial answer to these questions is: as long as liquidity suppliers have some constraints (for example due to inventory reasons or risk aversion), one can consider a model as in this paper, but where traders buy or sell at most n units of the asset, with n an appropriately large positive integer. The model must also account for traders who keep a permanent presence in the market. Such a model would no doubt be quite complicated.

Finally, one may worry about the assumption of independent Poisson arrivals. There is empirical evidence that arrivals are positively correlated: for example a market buy order is more likely to succeed another market buy order than any other type of orders

[12]There is evidence that market making arises endogenously in pure limit order markets. See for example Bloomfield, O'Hara, and Saar (2003).

(it is the diagonal effect in Biais et al. (1995)). Also, arrivals depend on bid and ask prices: the larger the bid-ask spread, the faster patient traders arrive at the market to supply liquidity. All these empirical facts can be accommodated into the model, but then the number of state variables would increase and the solution would become more difficult.

3. Equilibrium: One Side of the Book, Equal Arrival Rates

In this section, I analyze the sell-side of the limit order book, by assuming only two types of traders: patient sellers and impatient buyers. With the notation given above, $\lambda_{PB} = \lambda_{IS} = 0$. (By symmetry, one can derive similar results for the buy-side.) This case proves to be quite tractable. Moreover, it is also useful for understanding the general case, which can be thought as merging two one-sided models.

3.1. *Main Intuition*

Suppose the limit order book is empty, and a patient seller labeled "1" arrives first to 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.[13] Suppose 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 a limit order at $a_2 = A - \delta$ for some very small δ. 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 $a_2 < a_1 = A$ which is low enough. How is a_2 determined? By the condition that both traders have the same expected utility. 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

[13] I have 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 after relies on the infinite supply at A.

sellers, trader 1 has a limit order at $a_1 = A$, and trader 2 has a limit order at a_2. Of course, the values a_1 and a_2 are determined in equilibrium, and depend on what agents do in other states: imagine that instead of an impatient buyer who places a market order at a_2, there comes a patient seller who will place a limit order at a_3, and so on.

In solving for the equilibrium, it is surprising that allowing agents to freely cancel their limit orders, instead of complicating the solution actually simplifies it.[14] The main intuition is the following: in equilibrium, the existing sellers in the book compete for the incoming market orders from impatient buyers.[15] The sellers have their limit orders placed at different prices, but get the same expected utility: otherwise, they would "undercut by a penny" those with higher utility. Thus, the sellers with a higher limit order obtain in expectation a higher price, but also have to wait longer.[16] The fact that all sellers have the same expected utility makes the equilibrium Markov, and the number of the sellers in the book becomes 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 non-competitive 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. How can this equilibrium be sustained? By Nash threats: trader 1 can threaten 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. Non-competitive equilibria may be important for example in understanding dealer markets.[17] However, in the present paper I focus on competitive equilibria, since they are the more likely outcome of large, anonymous order-driven markets.

[14] In most financial markets cancellation of a limit order is free.

[15] The competition among patient traders does not drive expected profits down to zero in our model. Nor does this seem to happen in actual markets: Sandas (2001) shows that in the case of Stockholm Stock Exchange there are positive profits to be made by limit order traders.

[16] 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).

[17] 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.

3.2. *Description of the Equilibrium*

I assume a dynamic market clearing condition, namely that the arrival rates of patient sellers and impatient buyers are equal:

$$\lambda = \lambda_{PS} = \lambda_{IB} > 0.$$

The relevant parameters then are: A; B, which here is the reservation value of the sellers; r, the patience coefficient of the sellers; and λ.

I start with the easier problem of the existence of an equilibrium. In the end one wants the equilibrium to be competitive and stationary Markov, so one can directly look for an equilibrium where at each point in time the m sellers currently in the book have the same expected utility f_m. Since the state m follows a Markov process, f_m will satisfy a system of equations, which I call the "recursive system." Then I show that there exists a subgame perfect equilibrium by showing that there exists a solution to the recursive system. The strategies are easy to define, and it is also straightforward to show that the equilibrium is competitive, stationary, and Markov. The harder part is to show that the equilibrium above is unique in the class of rigid equilibria.

Before I begin a formal discussion of the results, I will give some intuition about how one searches for the equilibrium (proofs will be given later). First, the number of states must be finite. Denote by M the largest state. As m increases, each seller is strictly worse off, and the ask price will decrease. In state M it must be true that $f_M = B$, otherwise another patient seller would be tempted to join in. In this state, the bottom seller (the one with the lowest offer) has a mixed strategy: at the first arrival in an independent Poisson process with intensity μ, the seller will place a market order at B and exit.

Observe that from a state $m = 1, \dots, M-1$, the system can go either to: state $m-1$ if an impatient buyer arrives—after random time T_1; or to state $m+1$ if a patient seller arrives—after random time T_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 2λ (hence the expected value of T is $\frac{1}{2\lambda}$). Each event happens with probability $\frac{1}{2}$. One obtains the formula $f_m = \frac{1}{2}(f_{m-1} + f_{m+1}) - r \cdot \frac{1}{2\lambda}$. If one denotes by $\varepsilon = \frac{r}{\lambda}$, the formula becomes

$$2f_m + \varepsilon = f_{m-1} + f_{m+1}.$$

Similarly, from the terminal state M, the system can: go to $M-1$, if an impatient buyer arrives—after random time $T_1 \sim \exp(\lambda)$; stay at M, if a patient seller arrives—after random time $T_2 \sim \exp(\lambda)$; or go to $M-1$, if the current Bottom seller places a market order at B and exits—after random time $T_3 \sim \exp(\mu)$. 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. Then one gets the formula $f_M = f_{M-1} - r \cdot \frac{1}{\lambda+\mu}$. If one defines $u = \frac{\lambda}{\lambda+\mu} \in (0,1]$, the formula becomes

$$f_M + u\varepsilon = f_{M-1}.$$

Define also (this will be justified later): $f_0 = A$.

In conclusion, f_m satisfies a system of difference equations: the recursive system. How does one solve it? Since $f_0 = A$, the general solution is a quadratic function $f_m = A - bm + \frac{\varepsilon}{2}m^2$. To determine b and M, one uses $f_M = B$, together with $f_M + u\varepsilon = f_{M-1}$.

I now begin the formal discussion of the results. Start with some parameter values A, B, r, and λ. I define ε, M, u, μ, b as follows: First, let

$$\varepsilon = \frac{r}{\lambda}. \tag{1}$$

(It turns out that the properties of the equilibrium do not depend on r and λ individually, but only on their ratio ε.) For each real number u define

$$M_u = \tfrac{1}{2} - u + \sqrt{(\tfrac{1}{2} - u)^2 + \tfrac{2(A-B)}{\varepsilon}}.$$

One can easily show that M_u is decreasing in u, and that $M_0 - M_1 = 1$. Therefore, in the interval $[M_1, M_0)$ there is a unique integer M. This integer corresponds to a unique number $u \in (0,1]$ such that $M = M_u$. Finally, define

$$\mu = \frac{\lambda}{u} - \lambda, \tag{2}$$

$$b = \varepsilon(M - \tfrac{1}{2} + u). \tag{3}$$

DEFINITION 1. A sequence f_m, $m = 1, \ldots, M$, is called the *recursive system* associated to A, B, r and λ if for ε, M, and u defined as above the following formulas hold

$$\begin{cases} f_0 = A, \\ 2f_m + \varepsilon = f_{m+1} + f_{m-1}, \quad m = 1, \ldots, M-1 \\ f_M + u\varepsilon = f_{M-1}. \\ f_M = B. \end{cases} \tag{4}$$

PROPOSITION 1. *Given A, B, r, λ, there exists a unique solution f_m to the associated recursive system. It satisfies the formula*

$$f_m = A - bm + \frac{\varepsilon}{2}m^2, \quad m = 1, \ldots, M. \tag{5}$$

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: See Appendix A. □

I now state the main result of this section. Recall that a rigid equilibrium 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 (in this case, only by the seller with the lowest offer).

Theorem 2. *Given A, B, r, λ, there exists a subgame perfect equilibrium of the game: Let ε, M, f_m, μ be defined as above. Then in equilibrium there are at most M limit orders in the book, and the ask price in state $m = 1, \ldots, M$ is given by*

$$a_m = f_{m-1}, \quad \text{if } m < M; \tag{6}$$

$$a_M = B + \varepsilon. \tag{7}$$

The value function in state m is given by f_m. The strategy of each agent in state m 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. If not, then 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 after Poisson(μ) by placing a market order at B.*
- *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 m, and the ask price a_m. This equilibrium is unique in the class of rigid equilibria.

Proof: See Appendix A. □

Notice that there is some ambiguity in the way strategies are formulated, in the sense that, in state m, as long as some seller has a limit order at a_m, the other sellers can place their limit orders anywhere above a_m. This generates a whole class of equilibria, all of them equivalent from the Markov point of view: they lead to the same evolution of the state variables m and a_m. I consider the equilibria in this class to be the same equilibrium.

Let us give an example where strategies are completely defined: Suppose that a new seller arrives when there are already $m-1$ sellers in the book. Then I define the strategies by requiring 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, traders have their offers placed at $a_1, \ldots a_m$, and they never change them. One can think of this particular equilibrium as the one where there is an infinitesimal cost to modifying a limit order. Another example of an equilibrium is when in state m all sellers place their orders at a_m.

Also, by assumption one only considers equilibria in which strategies are mixed solely over time. If instead one considers mixing solely over actions, then there exists another competitive stationary Markov equilibrium. In that case, in state $M-1$ when the limit order book is about to become full, a newly arrived seller randomizes between entering the full state M (by placing a limit order), and not entering it (by placing a market order at B). The problem is that this equilibrium does not exist for all parameter values.

One final point is about the proof of uniqueness. To show that every rigid equilibrium is of the kind described in Theorem 2, one needs to know all the possible types of equilibrium behavior in the various states of the system. The most delicate situation is in state $m = M$, where one needs to decide which sellers can have mixed strategies. The more economically relevant choice seems to be the one in which only the bottom seller randomizes (this is the "rigid" equilibrium). In all the other cases the rest of the sellers have to randomize their orders, which seems less plausible. For more details, see the proof of the Theorem, together with Proposition 12 and Corollary 13 in Appendix A.

4. Equilibrium: One Side of the Book, Different Arrival Rates

In this section, I assume that the patient sellers arrive faster than the impatient buyers:

$$\lambda_1 = \lambda_{PS} > \lambda_2 = \lambda_{IB}.$$

There are two reasons why one might prefer this setup to the one where the arrival rates are equal. First, it corresponds to empirical evidence: Biais et al. (1995) show that limit orders arrive to the market faster than market orders. And second, the assumption $\lambda_1 > \lambda_2$ has the desirable implication that there is pressure for the spread to revert to its minimum value. This means that by using different arrival rates one can simulate a limit order book that has the property of *resiliency*.[18]

For this reason, it should be pointed out that the case of equal arrival rates is not a limiting case of different arrival rates. The two cases yield qualitatively different solutions. For example, the solution of the former case has an asymptotic limit when $\varepsilon \to 0$, while the solution of the latter does not.

4.1. *Description of the Equilibrium*

As in the case of equal arrival rates, I first give the intuition for the equilibrium, and then the formal results. Denote by f_m is the value function of a seller in the state where there are m sellers in the book. The number of states is finite, so there exists a largest state M. Moreover, $f_M = B$. From the state $m = 1, \ldots, M-1$ the the system can go to one of two states: $m+1$, if a patient seller arrives—after random time $T_1 \sim \exp(\lambda_1)$; or $m-1$, if an impatient buyer arrives—after random time $T_2 \sim \exp(\lambda_2)$. The arrival

[18]Foucault et al. (2004) define resiliency as the speed with which the bid-ask spread reverts to its minimum value. One can also simulate resiliency in my model by allowing the arrival rates of patient agents to depend on the spread, but this would lead to a more complicated model.

of the first of these two states happens at $\min(T_1, T_2) \sim \exp(\lambda_1 + \lambda_2)$. Then the first event happens with probability $\omega = \frac{\lambda_1}{\lambda_1+\lambda_2} > \frac{1}{2}$, and the second with probability $1 - \omega$. In one lets $\varepsilon = \frac{r}{\lambda_1+\lambda_2}$, one obtains

$$f_m + \varepsilon = \omega f_{m+1} + (1-\omega) f_{m-1}.$$

The characteristic equation $\omega x^2 - x + (1-\omega) = 0$ has two roots: $x_1 = 1$ and $x_2 = \alpha$, where $\alpha = \frac{1-\omega}{\omega} = \frac{\lambda_2}{\lambda_1} < 1$. Imposing the condition $f_0 = A$ as in the equal arrival case, the general solution is

$$f_m = A - C(1-\alpha^m) + \varepsilon\beta m, \qquad \text{with } C \text{ arbitrary and } \beta = \tfrac{1}{2\omega-1}.$$

The constant C is determined by looking at the final state M. In this state, suppose the bottom agent has the mixed strategy to place a market order at the first arrival in a Poisson process with intensity μ. Arrival of patient seller does not matter, so one gets the equation $(\lambda_2 + \mu) f_M + \varepsilon = (\lambda_2 + \mu) f_{M-1}$. Define $u = 1/(1 - \omega + \frac{\mu}{\lambda_1+\lambda_2})$. Then one gets

$$f_M + u\varepsilon = f_{M-1}.$$

One now uses this equation together with $f_M = B$ to determine M and C.

I now begin the formal discussion of the results. Start with some parameter values A, B, r, λ_1, λ_2. I define ε, ω, α, β, M, u, μ, and C as follows: First, let

$$\varepsilon = \frac{r}{\lambda_1 + \lambda_2}, \quad \omega = \frac{\lambda_1}{\lambda_1 + \lambda_2} > \frac{1}{2}. \tag{8}$$

$$\alpha = \frac{1-\omega}{\omega} < 1, \quad \beta = \frac{1}{2\omega - 1}. \tag{9}$$

The integer $M > 0$ and $u \in (0, \frac{1}{1-\omega}]$ are defined to be the unique ones for which

$$A - B + \varepsilon\beta M = \frac{\varepsilon(\beta + u)}{\alpha^{M-1} - \alpha^M}(1 - \alpha^M). \tag{10}$$

Define

$$C = \frac{\varepsilon(\beta + u)}{\alpha^{M-1} - \alpha^M}, \quad \mu = \frac{1 - u(1-\omega)}{u}(\lambda_1 + \lambda_2). \tag{11}$$

PROPOSITION 3. *Given A, B, r, λ_1, λ_2, there exists a unique solution f_m to the associated recursive system*

$$
\begin{cases}
f_0 = A, \\
f_m + \varepsilon = \omega f_{m+1} + (1-\omega) f_{m-1}, \quad m = 1, \ldots, M-1 \\
f_M + u\varepsilon = f_{M-1}, \\
f_M = B.
\end{cases} \tag{12}
$$

The solution f_m is given by

$$
f_m = A - C(1 - \alpha^m) + \varepsilon\beta m, \quad m = 1, \ldots, M. \tag{13}
$$

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 $m > M$.

The proof of the next result is essentially identical to that of Theorem 2.

Theorem 4. *Given $A, B, r, \lambda_1 > \lambda_2$, there exists a competitive stationary Markov equilibrium of the game. Let $\varepsilon, \omega, M, f_m$ be as defined above. Then in equilibrium there are at most M limit orders in the book, and the ask price in state $m = 1, \ldots, M$ is given by*

$$
a_m = f_{m-1}, \quad \text{if } m < M; \tag{14}
$$

$$
a_M = B + (1-\omega)\varepsilon. \tag{15}
$$

The value function in state m is given by f_m. The strategies of agents are the same as in Theorem 2. This equilibrium is unique in the class of rigid equilibria.

REMARK 1. An interesting feature of these formulas when $\omega > \frac{1}{2}$ is that one can calculate the limit of equation (13) when $\varepsilon \to 0$. To see why, notice that $f_M = B$, so $C(1 - \alpha^M) = A - B + \varepsilon\beta M$. The number α^M is of the order of ε, so M is of the order of $\log \frac{1}{\varepsilon}$. (See the proof of Proposition 5 for more details.) This implies that $C = A - B$ modulo terms of order $\varepsilon \log \frac{1}{\varepsilon}$, which is of smaller order for example than $\varepsilon^{1/2}$. In the end, one gets the following formula

$$
f_m \approx B + (A - B)\alpha^m \quad \text{if } \varepsilon \approx 0. \tag{16}
$$

By contrast, in the case of equal arrival rates ($\omega = \frac{1}{2}$), there is no limit of equation (5) when $\varepsilon \to 0$. Instead, after some rescaling, the whole recursive system has an asymptotic limit (this is true in the general case: see Theorem 9). The limiting f is a function defined on $[0, \gamma]$, and it satisfies the following differential equation:

$$\begin{cases} f'' = 1, \text{ such that} \\ f(0) = A, \\ f'(\gamma) = 0, \\ f(\gamma) = B. \end{cases} \tag{17}$$

(The free boundary point γ corresponds to M after the appropriate rescaling.)

4.2. *The Distribution of the Bid-Ask Spread*

Once the equilibrium evolution of the limit order book is known, it is interesting to look at the properties of the bid and ask prices, and in particular at the distribution of the bid-ask spread. What are the determinants of the average bid-ask spread? In classical market microstructure, this is determined by the amount of asymmetric information in the market. In this paper, the spread is determined by how fast traders arrive at the market. More precisely, recall that λ_1, λ_2 are the arrival rates of patient and impatient traders, respectively; and $\omega = \frac{\lambda_1}{\lambda_1+\lambda_2} > \frac{1}{2}$ represents the percentage of patient trader arrivals of the total population arrivals. Define

$$\varepsilon = \frac{r}{\lambda_1 + \lambda_2}, \quad \beta = \frac{1}{2\omega - 1}, \quad g = \frac{\omega}{1-\omega}.$$

Proposition 5. *In the context of Theorem 4, let $\bar{S}$ be the average equilibrium bid-ask spread. Then when ε is small, $\bar{S}$ can be approximated by*

$$\bar{S} \approx \beta \frac{g}{\log g} \varepsilon \log \frac{1}{\varepsilon}. \tag{18}$$

As a consequence, asymptotically $\bar{S}$ is of the order of $\varepsilon \log \frac{1}{\varepsilon}$.

Proof: See Appendix A. □

This is to be compared with Farmer, Patelli and Zovko (2003), who in their cross-sectional analysis of the London Stock Exchange show that with a high R^2 the average bid-ask spread varies as $\varepsilon^{3/4}$. It is interesting that Farmer et al. arrive to their formula not from a theoretical model, but from an exogenous statistical model of market and limit order arrivals. Their ε equals the average market order size divided by the arrival rate of market orders (and normalized by the rate of cancellation δ of limit orders). If in my model I take a unit to be the average market order size, then the two parameters ε are identical up to the patience coefficient r (normalized by the cancellation rate δ). It is assumed that r is the same across stocks.

5. Multi-Unit Market Orders and Price Impact

In the previous section, one saw that when agents have one-unit demands the only important limit order in equilibrium is the one at the ask, and the other limit orders can be at any level above the ask. So the only results that the one-unit framework can tackle are for the bid and ask prices. But to determine the distribution of the other limit orders one must consider the possibility of multi-unit orders. Let k be the maximum number of units that a market order can have with positive probability. Then the equilibrium offers in the book will be fixed up to k levels starting with the ask.

In the rest of the paper, the following notions are synonymous: price impact function, configuration of limit orders, and shape of the limit order book: Given the configuration of limit orders, i.e. the levels at which the limit orders are placed at any point in time, one can define the (instantaneous) price impact function *Imp* as follows: for each integer $i > 0$ define by $Imp(i) = a_{i+1} - a_1$, where a_{i+1} is level of the $i'th$ offer above the ask. The price impact represents how much the price moves against a trader who submits an i-unit market order.

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. In this paper, since there is no tick size, prices can be any real number. So, by convention, depth at a certain discrete value can be obtained by collecting all the one-unit limit orders around the corresponding value. Then what in the literature is called a hump-shaped limit order book in our model is translated by the fact that limit orders cluster at some point above the ask. In that case depth becomes larger around that point, thus creating a hump in the graph.

The intuition why sellers would want to cluster above the ask is simple: they want to take advantage in the best way possible of the incoming multi-unit market orders. This generates an equilibrium shape of the limit order book, which I analyze next.

5.1. *Description of the Equilibrium*

Let k be the maximum number of units that an order can have with positive probability. Assume that patient sellers still arrive with only one unit to sell. Define

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

Assume that $\lambda_i > 0$ for all $i = 1, \dots, k$. Moreover, as in the previous section, one wants the sellers to arrive faster than the units demanded by the buyers. This is equivalent to

$$\lambda > \sum_{i=1}^{k} i\, \lambda_i. \tag{19}$$

As before, the number of states is finite, so there exists a largest state M. Moreover, $f_M = B$. From the state $m = 1, \dots, M-1$ the the system can go to one of the following states: $m+1$, if a patient seller arrives; or $m-i$, $i = 1, \dots, k$ if an impatient i-buyer arrives. In state M there is some randomization: the bottom seller may leave after the first arrival of a Poisson process with intensity μ.

The recursive system then takes the following form:

$$\begin{cases} f_0 = f_{-1} = \cdots = f_{1-k} = A, \\ (\lambda + \sum_{i=1}^{k} \lambda_i) f_m + r = \lambda f_{m+1} + \sum_{i=1}^{k} \lambda_i f_{m-i}, \\ (\sum_{i=1}^{k} \lambda_i + \mu) f_M + r = (\lambda_1 + \mu) f_{M-1} + \sum_{i=2}^{k} \lambda_i f_{M-i}. \\ f_M = B. \end{cases} \tag{20}$$

Similar to the results of Section 4, one obtains the following result:

Proposition 6. *The solution of the recursive system takes the following form*

$$f_m = C_0 + C_1 \alpha_1^m + C_2 \alpha_2^m + \cdots + C_k \alpha_k^m + \varepsilon\, m, \qquad \textit{where} \tag{21}$$

$$\varepsilon = \frac{r}{\lambda - \sum_{i=1}^{k} i \lambda_i} > 0 \qquad \textit{and} \qquad |\alpha_1|, |\alpha_2|, \dots, |\alpha_k| < 1. \tag{22}$$

The complex numbers: $\alpha_0 = 1, \alpha_1, \ldots, \alpha_k$ *are the roots of the polynomial*

$$P(X) = \lambda X^{k+1} - (\lambda + \sum_{i=1}^{k} \lambda_i) X^k + \sum_{i=1}^{k} \lambda_i X^{k-i}. \tag{23}$$

The description of the equilibrium is the following:

Theorem 7. *Given* A, B, r, λ *and* $\lambda_i, i = 1, \ldots, k$ *which satisfy the inequalities above, there exists a competitive stationary Markov equilibrium of the game. Denote by* $i_0 = \min\{k, m\}$*. In equilibrium there are at most* M *limit orders in the book, and if* $i = 1, \ldots, i_0$*, then the level of the i'th limit order (counted from bottom up) in state* $m < M$ *is given by*

$$a_i(m) = \frac{\lambda_k f_{m-k} + \lambda_{k-1} f_{m-k+1} + \ldots + \lambda_i f_{m-i}}{\lambda_k + \lambda_{k-1} + \ldots + \lambda_i}, \tag{24}$$

where by convention $f_0 = f_{-1} = \cdots = f_{1-k} = A$*. The value function in state* m *is given by* f_m*. 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_M*, who exits (by placing a market order at* B*) after the first arrival in a Poisson process with intensity* μ*.*
- *If* $m > M$*, then immediately place a market order at* B*.*

This equilibrium is unique in the class of rigid equilibria.

One can make these formulas more explicit. There are two cases, depending on whether the k-unit market orders clear all the limit orders in the book or not.

Case 1: $m \geq k$.

$$\begin{aligned}
a_k &= f_{m-k}, \\
a_{k-1} &= \frac{\lambda_k f_{m-k} + \lambda_{k-1} f_{m-k+1}}{\lambda_k + \lambda_{k-1}}, \\
a_{k-2} &= \frac{\lambda_k f_{m-k} + \lambda_{k-1} f_{m-k+1} + \lambda_{k-2} f_{m-k+2}}{\lambda_k + \lambda_{k-1} + \lambda_{k-2}}, \\
&\vdots \\
a_1 &= \frac{\lambda_k f_{m-k} + \lambda_{k-1} f_{m-k+1} + \cdots + \lambda_1 f_{m-1}}{\lambda_k + \lambda_{k-1} + \cdots + \lambda_1},
\end{aligned}$$

Since it does not matter what happens above a_k, one can also choose by convention: $a_{k+1} = f_{m-k-1}, \ldots, a_{m-1} = f_1, a_m = f_0$.

Case 2: $m < k$.

$$\begin{aligned}
a_m &= \frac{\sum_{i=m}^k \lambda_i A}{\sum_{i=m}^k \lambda_i} = A, \\
a_{m-1} &= \frac{\sum_{i=m}^k \lambda_i A + \lambda_{m-1} f_1}{\sum_{i=m}^k \lambda_i + \lambda_{m-1}}, \\
&\vdots \\
a_1 &= \frac{\sum_{i=m}^k \lambda_i A + \lambda_{m-1} f_1 + \cdots + \lambda_1 f_{m-1}}{\sum_{i=m}^k \lambda_i + \lambda_{m-1} + \cdots + \lambda_1}.
\end{aligned}$$

5.2. *Price Impact of Transactions: Numeric Results*

As pointed out before, in the case of multi-unit market orders it may be optimal for agents to cluster away from the ask. What determines the shape of the limit order book are the values $\lambda, \lambda_1, \ldots, \lambda_k$, which indicate how likely it is for an i-unit market order to arrive.[19]

When one analyzes the shape of the price impact function, it turns out that a crucial factor is how fast the rates λ_i decrease. To get some intuition, suppose that λ_1 is larger than the rest. For example, consider a function $\phi(i)$ which is decreasing in i, and some

[19] One can argue in fact that what matters here are the *expectations* that traders have about the arrival rates of the incoming market orders, and not the actual values. But if one assumes that traders have rational expectations, those values should be the same.

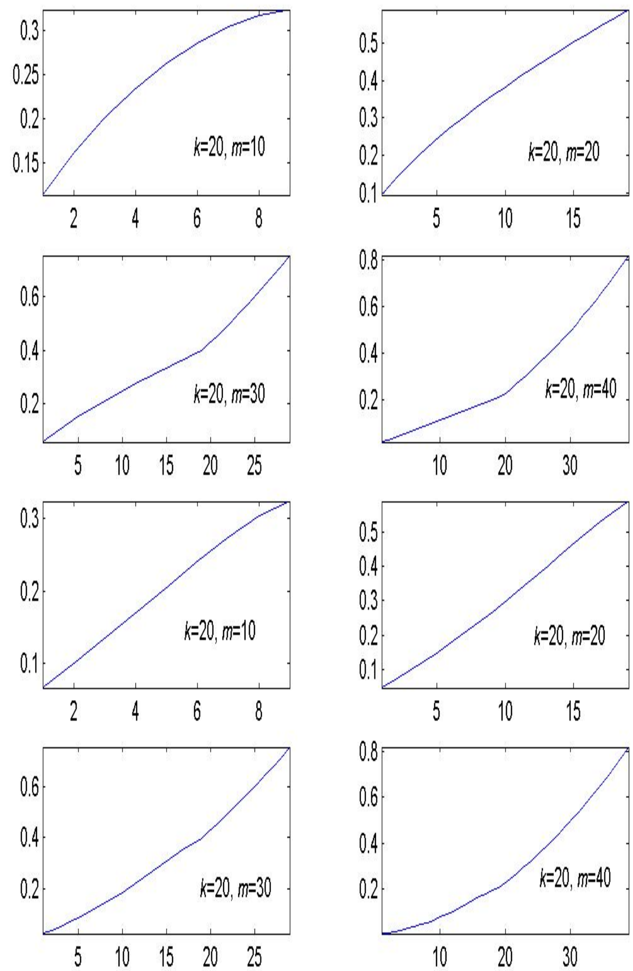

FIGURE 1. The instantaneous price impact function $Imp(i) = a_{i+1} - a_1$ plotted against i. ($Imp(i)$ is the difference in the level of the $i+1$'st limit sell order above the ask and the ask price.) The values of parameters are $A = 1, r = 0.001, \delta = 0.04, k = 20$, and the arrival rates are $\lambda_1 = 1$; $\lambda_i = \lambda_0 \phi(i), i = 2, \ldots, k$, where $\lambda_0 = 10^{-5}$; $\lambda = \sum_{i=1}^{k} i\lambda_i$. The top four plots correspond to the weight function $\phi(j) = 1/j(j+1)$; the bottom four plots to $\phi(j) = 4/2^j$. Each set of four graphs is considered for the case when there are $m = 10, 20, 30, 40$ sell orders in the book. The number k is the maximum number of units that market buy orders can have, and $\phi(i)$ indicates how fast the arrival rate of i-market orders decreases with i. For the top four plots (when $\phi(i)$ does not decrease too fast), notice the concave shape of the price impact function $Imp(i)$ in the region where $i \leq k$. For the other regions notice either linearity or convexity of price impact.

small value $\lambda_0 > 0$ such that

$$\lambda_1 = 1 \quad \text{and} \quad \lambda_i = \lambda_0 \phi(i), \text{ if } i \geq 2.$$

Then set $\lambda = \sum_{i=1}^{k} i\lambda_i$, plus some small number so that one has indeed $\lambda > \sum_{i=1}^{k} i\,\lambda_i$ (in fact it can be equal, it does not change the analysis). Now calculate the price impact function for different choices of $\phi(i)$, $i = 2, \ldots, k$. Take for example

$$\phi_1(i) = \frac{1}{i(i+1)}, \qquad \phi_2(i) = \frac{4}{2^i}.$$

The price impact function in state m is defined as the change in the ask price when i units are bought via a market order of i units:

$$Imp(i, m) = a_{i+1}(m) - a_1(m), \quad \text{as a function of } i \text{ (and } m\text{)}. \tag{25}$$

To calculate the price impact function, one can apply the theoretical results of the previous section in the following way: Equation (20) shows that f_m satisfies a recursive formula with initial conditions $f_0 = f_{-1} = f_{1-k} = A$. This fixes k coefficients of f_m in equation (21). To fix the last coefficient, one should use the equations in (20). For simplicity, I choose a different method: Let B become a free parameter, and choose another parameter $\delta > 0$, so that

$$f_1 = A - \delta.$$

This fixes all coefficients C_i, so one can determine M as the first m for which f_m starts increasing (according to Proposition 6). Then one determines B by $B = f_M$. So in the analysis that follows and in Figure 1, instead of B, I use the parameter δ.

In Figure 1, I compare the graphs of the price impact function $Imp(i, m)$ for the two functions $\phi_1 = 1/i(i+1)$ and $\phi_2 = 4/2^i$. In both cases the maximum number of sellers is 41 (so $m < 41$), and I display the results when the book has $m = 10, 20, 30, 40$ limit sell orders. All the graphs shown are for $k = 20$, which means that market buy orders for more than 20 units appear with zero probability. In the first case (for ϕ_1), the arrival rate λ_i of i-market orders does not decrease very fast in i, while in the second case the arrival rate decreases exponentially. The top four plots refer to ϕ_1, and the bottom four to ϕ_2.

Notice that in the case of ϕ_1, the price impact function $Imp(i)$ is concave when $i \leq k$ units are purchased (recall that k is the maximum size of a market buy order that sellers in the book expect to occur). So when m itself is less than k, the price impact function is concave everywhere, which is the case of the first two graphs. The intuition for this finding is the following: each seller above the ask up to level k expects that his limit order will be cleared by a market order with a probability which is not too small. Then instead of clustering near the ask, they prefer to take advantage of the large market orders and cluster above the ask. This leads to a concave price impact function. Above the level k or when the probability of a large market order decreases too fast, the price impact function is linear, and even convex. This is the case for all the graphs for ϕ_2, and in the region where $i > k$ for ϕ_1.

Overall, the conclusion seems to be that for smaller orders the price impact function should be mildly concave, and for larger orders it should be mildly convex.[20] This reflects the existing differences of opinions 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.

6. Equilibrium: The General Case

Consider the general case, when all types of buyers and sellers arrive with positive probability. For simplicity, I assume that all the arrival rates are equal:

$$\lambda = \lambda_{PB} = \lambda_{PS} = \lambda_{IB} = \lambda_{IS} > 0.$$

Later on, I indicate what happens when the arrival rates are different. As in the one-sided case, the most important situation is when $\lambda_1 = \lambda_{PB} = \lambda_{PS} > \lambda_2 = \lambda_{IB} = \lambda_{IS}$.

To get some intuition about the equilibrium, consider a setup similar to that of the one-sided case, but suppose that after a patient seller (which has a limit sell order at A) a patient buyer arrives at the market. Then the buyer behaves as a monopolist towards the potential incoming impatient sellers, and places a limit buy order at B. In this

[20]The exact predictions of the model are actually slightly different: as soon as the order reaches a certain size (equal to the existing depth in the limit 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 step in to supply liquidity at those prices.

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 like in the one-sided case, where new patient agents just keep placing bid-ask improving limit orders until it is better to trade immediately rather than wait. Thus, patient agents form two queues, a descending one starting from A, and an ascending one starting from B.

What happens in a state where the limit order book is full? Then the traders on both sides play a game of attrition. As before, a *rigid* equilibrium is a competitive stationary Markov equilibrium in which the 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. I call such a limit order *fleeting*. In the state where a fleeting order is placed all traders have the same expected utility h.

Unlike the previous sections, in the general case even after restricting attention to rigid equilibria, one still does not get uniqueness. Nevertheless, all rigid equilibria are close to each other, in a sense that will be made precise below.

6.1. *Description of the Equilibrium*

I first give some intuition for the definitions made in this section, and then prove the formal results. In the state with m sellers and n buyers, denote by $a_{m,n}$ the ask price, $b_{m,n}$ the bid price, $f_{m,n}$ the expected utility of the sellers, and $g_{m,n}$ (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 where at least some agent has a mixed strategy. The role played in the one-sided case by the states $m = 1, \ldots, M$ is here played by Ω, while the role of M is played by the boundary γ of the state region.

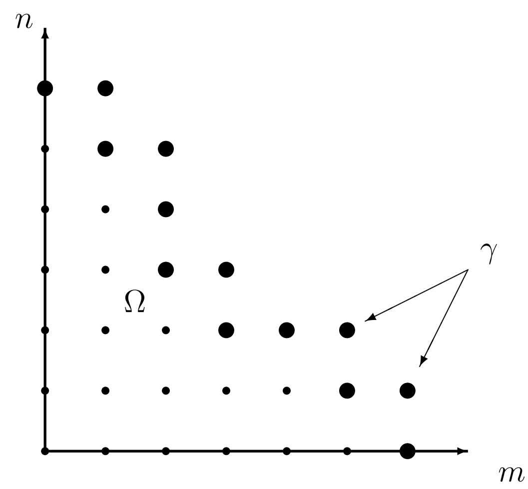

FIGURE 2. The state region Ω and the boundary γ.

I conjecture that Ω is such that if (m, n) belongs to Ω, then also $(m-1, n)$ and $(m, n-1)$ are in Ω (as long as the coordinates are non-negative). This will be justified later in the discussion of uniqueness, but for now I just assume it. Moreover, I assume that on each 45-degree line in the first quadrant that intersects Ω there exists a unique point in γ. An example of such a state region is given in Figure 2. This allows one to define various types of points in Ω, as in Figure 3.[21]

As in the one-sided case, it is a good idea to find a recursive structure for the value 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 places 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 places a market order at B (when $m = 0$);
- $(m, n+1)$, if a patient buyer arrives and submits a limit order;

[21]There are two more type of boundary points in Ω. For example, in Figure 3 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, I decided to ignore these types of points.

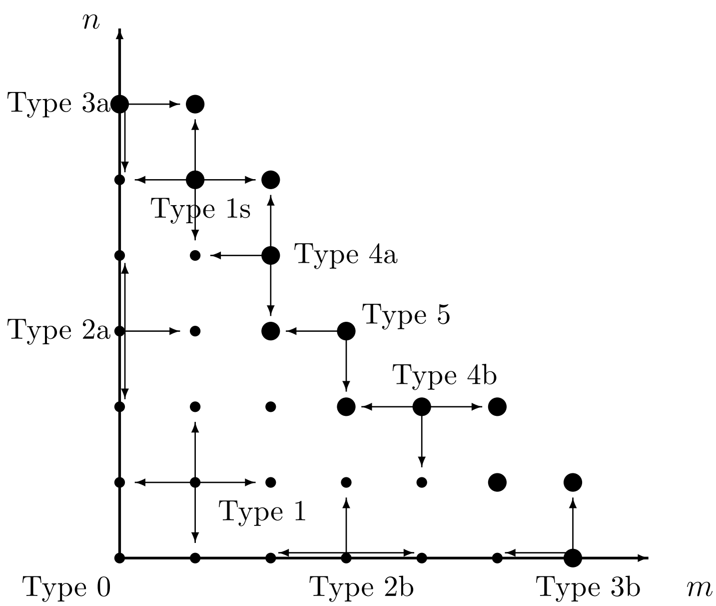

FIGURE 3. Types of points in the state region Ω.

- $(m-1, n-1)$, if after a positive expected time a seller places a fleeting limit order and a buyer immediately accepts.

From a state (m, n) of Type 1 the system can go only to the states $(m-1, n)$, $(m+1, n)$, $(m, n-1)$, or $(m, n+1)$. The arrival of the first of these four states happens after a random time, which is exponentially distributed with parameter 4λ. Then each event happens with probability $1/4$. One obtains the formula $f_{m,n} = \frac{1}{4}(f_{m-1,n} + f_{m+1,n} + f_{m,n-1} + f_{m,n+1}) - r \cdot \frac{1}{4\lambda}$. If one denotes by $\varepsilon = \frac{r}{\lambda}$., the formula becomes

$$4f_{m,n} + \varepsilon = f_{m-1,n} + f_{m+1,n} + f_{m,n-1} + f_{m,n+1}.$$

Take now a point of Type 1s. In this case, after a random time $T \sim \exp(\mu)$, the bottom seller submits a fleeting limit order at $h = f_{m-1,n-1} = g_{m-1,n-1}$, and the top buyer immediately accepts it. If one denotes $s = \mu/\lambda$, one obtains the equation

$$(4+s)f_{m,n} + \varepsilon = f_{m-1,n} + f_{m+1,n} + f_{m,n-1} + f_{m,n+1} + sf_{m-1,n-1}.$$

This type of reasoning works for all the other states in Ω, which gives a recursive system that one needs to solve. It is not hard to show that the solution to the recursive system yields a competitive stationary Markov equilibrium.

I now start discussing the formal results. The relevant parameters are A, B, r, and λ. Define

$$\varepsilon = r/\lambda. \tag{26}$$

Definition 2. Consider a region Ω in the positive quadrant which satisfies the property: 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 the each boundary point $(m,n) \in \gamma$, consider a number $s_{m,n} \geq 0$. Let s be the collection of all $s_{m,n}$. Then I define the recursive system associated to (Ω, s) 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 1s,

$$(4 + s_{m,n})f_{m,n} + \varepsilon = f_{m-1,n} + f_{m+1,n} + f_{m,n-1} + f_{m,n+1} + s_{m,n}f_{m-1,n-1},$$
$$(4 + s_{m,n})g_{m,n} - \varepsilon = g_{m-1,n} + g_{m+1,n} + g_{m,n-1} + g_{m,n+1} + s_{m,n}g_{m-1,n-1},$$
$$f_{m,n} = g_{m,n};$$

- 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,

$$f_{0,n} = A,$$
$$(2 + s_{0,n})g_{0,n} - \varepsilon = (1 + s_{0,n})g_{0,n-1} + g_{1,n},$$
$$f_{0,n} = g_{0,n};$$

and similarly for Type 3b.

- If (m,n) is of Type 4a,

$$(4 + s_{m,n})f_{m,n} + \varepsilon = f_{m-1,n} + 2f_{m,n-1} + f_{m,n+1} + s_{m,n}f_{m-1,n-1},$$
$$(4 + s_{m,n})g_{m,n} - \varepsilon = g_{m-1,n} + 2g_{m,n-1} + g_{m,n+1} + s_{m,n}g_{m-1,n-1},$$
$$f_{m,n} = g_{m,n};$$

and similarly for Type 4b.

- If (m,n) is of Type 5,

$$(4 + s_{m,n})f_{m,n} + \varepsilon = 2f_{m-1,n} + 2f_{m,n-1} + s_{m,n}f_{m-1,n-1},$$
$$(4 + s_{m,n})g_{m,n} - \varepsilon = 2g_{m-1,n} + 2g_{m,n-1} + s_{m,n}g_{m-1,n-1},$$
$$f_{m,n} = g_{m,n};$$

If (m,n) is not in Ω, and $m,n > 0$, consider the unique point (m',n') in γ that lies on the 45-degree 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-degree 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, respectively. Finally, if (m,n) is not in Ω, and $n = 0$, define $f_{m,n} = g_{m,n} = B$; and similarly when $m = 0$.

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 $s_{m,n} \geq 0$,

$$a_{m,n} = f_{m-1,n} + s_{m,n}(f_{m-1,n-1} - f_{m,n}),$$

$$b_{m,n} = g_{m,n-1} + s_{m,n}(g_{m-1,n-1} - g_{m,n});$$

the formulas for the other types of boundary points are similar.

The main result of this section is that, given an appropriate solution of the recursive system, there exists an equilibrium of the game.

Theorem 8. *Given A, B, r, and λ, suppose $f_{m,n}$ and $g_{m,n}$ are a solution of the recursive system associated to a pair (Ω, s) as in Definition 2. Consider also the associated set of numbers $a_{m,n}$ and $b_{m,n}$ as defined above, and assume that $A \geq a_{m,n} \geq b_{m,n} \geq B$ for all $(m,n) \in \Omega$. Then there exists a competitive stationary Markov equilibrium of the game, which is also rigid. For this equilibrium, the state region is Ω, and the boundary of the state region is γ. The state variables are: m, n, the ask price $a_{m,n}$, and the bid price $b_{m,n}$. To describe the equilibrium strategies, let (m,n) be a state, not necessarily in Ω. As before, one only needs to describe 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.)*

- *If (m,n) is in Ω, but not in 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*, let* $\mu_{m,n} = \lambda s_{m,n}$*. Then the strategy is the same as the one above, except that after Poisson(*$\mu_{m,n}$*) 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: See Appendix A. □

The theorem only guarantees the existence of an equilibrium. Uniqueness is a much more delicate matter. As a first step, one would like to show that for any rigid equilibrium the state region Ω has a nice shape, i.e., that all interior points (surrounded by γ and the coordinate axes) belong to Ω. This would be true if one could show the following property: if (m, n) is in Ω, then $(m - 1, n)$ and $(m, n - 1)$ are also in Ω. This in turn would be easy to show if one could prove the following intuitive conjecture.

Conjecture 1. In any rigid equilibrium, the arrival of a new seller at a state in Ω makes the sellers worse off and the buyers better off. Moreover, the sellers are worse off by more than the buyers are better off.

The author was not able to prove this result, although numerically it verified in all the cases that were studied. The conjecture can be justified heuristically in various ways. One is given in Remark 2 of Appendix A.

But even if the conjecture were true, for some values of the parameters A, B, r, λ it turns out that there are several distinct pairs (Ω, s) for which there exists a solution to the associated recursive system. So uniqueness in the strict sense fails, and one can only hope that the solutions must be close to each other in some appropriately defined way. And indeed, one can show that asymptotically the solution is unique. The recursive system in Definition 2 that $f_{m,n}$ and $g_{m,n}$ satisfy is formed of finite difference equations, so when ε is small one may expect f and g to approach the solution of a system of differential equations. To see that this is true, let $\varepsilon = \delta^2$, $x = m\delta$ and $y = n\delta$. Define the functions f and g at the discrete values $(x, y) = (m\delta, n\delta)$ by

$$f(x, y) = f_{m,n}, \qquad g(x, y) = g_{m,n}.$$

Then one obtains the following asymptotic result:

Theorem 9. *Any solution of the recursive system in Definition 2 converges when $\varepsilon = r/\lambda$ is small to the solution of the following system of partial differential equations with a free boundary γ:*

$$\begin{cases} \triangle f = 1, \\ f(0, y) = A, \\ \frac{\partial f}{\partial y}(x, 0) = 0, \\ \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} = 0 \ \textit{at}\ \gamma; \end{cases} \qquad \begin{cases} \triangle g = -1, \\ g(x, 0) = B, \\ \frac{\partial g}{\partial x}(0, y) = 0, \\ \frac{\partial g}{\partial x} + \frac{\partial g}{\partial y} = 0 \ \textit{at}\ \gamma; \end{cases} \tag{27}$$

where the free boundary γ is determined by the condition

$$f = g \ \textit{at}\ \gamma. \tag{28}$$

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

Proof: See Appendix A. □

Each partial differential equations 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 for example Gladwell and Wait (1979).

I close this section with a brief discussion about different arrival rates. As pointed out in the one-sided case, this is an important case, since it leads to a resilient limit order book. So consider the case when $\lambda_1 = \lambda_{PB} = \lambda_{PS} > \lambda_2 = \lambda_{IB} = \lambda_{IS}$. As in Section 4, define

$$\varepsilon = \frac{r}{\lambda_1 + \lambda_2}, \quad \omega = \frac{\lambda_1}{\lambda_1 + \lambda_2} > \frac{1}{2}, \quad \alpha = \frac{1 - \omega}{\omega} = \frac{\lambda_2}{\lambda_1} < 1.$$

Denote by $f_{m,n}$ and $g_{m,n}$ the solution to the corresponding recursive system as in Definition 2. One can easily check that when ε is small, f, g, and γ have the following limiting behavior:

$$\left\{\begin{array}{l} f_{m,n} = B + (A-B)\alpha^m, \\ g_{m,n} = A - (A-B)\alpha^n, \\ \gamma: \quad \alpha^m + \alpha^n = 1. \end{array}\right. \tag{29}$$

(Also, Ω here is the set of states for which $\alpha^m + \alpha^n \geq 1$.) The intuition for this result is that in the limit the sell-side and the buy-side of the book become decoupled, and each value function satisfies the formula corresponding to the one-sided limit order book, as in Equation (16). Notice that when $\alpha < 1/2$, the state region Ω contains only states with either $m = 0$ or $n = 0$. This is essentially the result of Section 7, which represents the limiting case $\alpha = 0$ when no impatient traders arrive, and all the agents are equally patient.

6.2. *Numeric Results*

Having described the equilibrium based on the pair (Ω, s), one may wonder how to actually find such a pair starting from A, B, and ε. This is far from trivial. Of course, one could take the brute force approach and for each Ω within reasonable values try to solve the system. But for ε small the complexity of this approach becomes daunting. The reason is that this is a system of equations with a free boundary γ, and one needs to simultaneously find the solution of the system and the shape of the boundary.

To find the state region Ω, one uses the intuition coming from the asymptotic result in Theorem 9. One sees that the asymptotic boundary is slightly concave, so it is a good idea to try regions state regions Ω which are close to being triangular (bounded by the coordinate axes and the line $X + Y = R$). As one increases the size of Ω, one will be forced to take out a few points from the triangle; otherwise, there would be no solution to the system of equations. Indeed, in the examples I computed, the shape of Ω is close to being triangular only when ε is relatively large. As ε gets smaller, points on and below the diagonal $X + Y = R$ start disappearing from Ω (see Table 1).

TABLE 1. Solution in the general case with both buyers and sellers, for $A = 1, B = 0, \varepsilon = 0.09$. Left bottom corner corresponds to state $(0, 0)$. The number in position (m, n) represents the value function $f_{m,n}$ for the sellers in state (m, n). The state region Ω is the same as in Figure 2. The vector s collects the variables corresponding to the mixed strategies along the boundary γ, starting from $(0, 6)$ down to $(6, 0)$ along γ. The value function $g_{m,n}$ for 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 one.

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

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

When patient traders arrive faster than impatient traders, so $\lambda_{PS} = \lambda_{PB} > \lambda_{IS} = \lambda_{IB}$, the waiting costs of patient agents increase. That implies that the limit order book would be more "rarefied" than in the case when all arrival rates are equal. The first guess is that the regions Ω for which one can find solutions are more convex than when all arrival rates are equal. Numeric experiments show this to be the case. This is consistent with the limiting case when only patient traders arrive at the market. In Section 7, I show that indeed the region Ω becomes so convex that it collapses to the coordinate axes.

6.3. *Empirical Implication: Market Orders and the Spread*

An implication of the equilibrium in the general case is that a market sell order leads to an decrease in both the bid and the ask, but the decrease in the bid is larger. The intuition for this was stated before: the departure of one limit buyer makes the buyers better off, and the sellers worse off (because of future possibility of trading with each other when the book becomes full). But that possibility is remote enough so that the decrease in the value function of the sellers is less than the decrease in the value function of the buyers (recall that the value function is minus the utility for the buyers).

This implication was noticed by Biais, Hillion and Spatt (1995), in their analysis of the order flow in the Paris Bourse. Their analysis goes as follows: *"The downward shift in the book has two components, the decrease in the bid, merely reflecting that the large sale consumed the liquidity offered at that quote, and the subsequent decrease in the ask, reflecting the reaction of the market participants to the large sale. The decrease in the bid could be a transient decrease in the liquidity on this side of the book, or a permanent information adjustment. In our one-lag analysis, we cannot differentiate the two hypotheses. In contrast to the behavior of the bid, the decrease in the ask is likely to reflect the information effect."*

I argue that the decrease in the ask need not reflect an information effect. Indeed, it can simply be regarded as an adjustment made by the limit sellers, who, after the bid decreased, realize that they now have to wait more to execute their orders, and lower their offers accordingly.

7. Equilibrium: The Homogeneous Case

I now study the case when all agents are equally patient: $\lambda_{IS} = \lambda_{IB} = 0$. For simplicity, I also assume that the arrival rates of patient sellers and patient buyers are equal:

$$\lambda = \lambda_{PS} = \lambda_{PB} > 0.$$

It turns out that in this case the limit order book cannot accommodate both buyers and sellers. This is because in general patient traders extract rents from the impatient ones. But when all agents are equally patient, they cannot extract rents from each other, so a bargaining game follows. In principle, there can be many equilibria, unless one puts more structure on the bargaining problem. I do not follow this path here. Instead, I

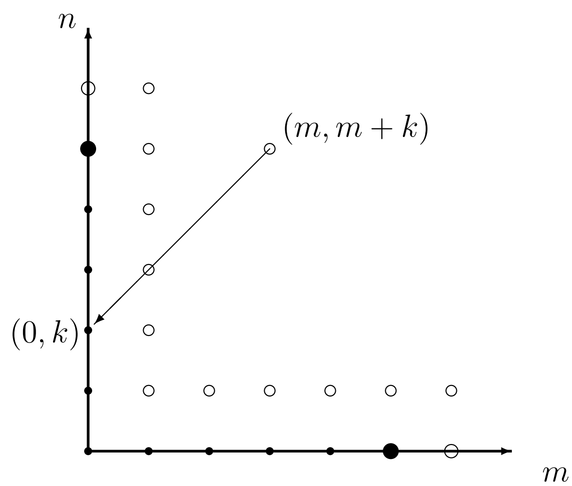

FIGURE 4. The state region in the homogeneous case (surrounding states are marked with a circle).

give an example of a competitive stationary Markov equilibrium, which seems to be the prototype for all such equilibria.

In this equilibrium, there are either only sellers or only buyers present in the limit order book. The case when there are only sellers in the book is very similar to the one-sided case (with only patient sellers and impatient buyers). In that case, the sellers behave like patient sellers: they place their limit orders and wait, while the buyers behave like impatient buyers: when they arrive they immediately place market orders and exit the game. The interesting difference is that the first seller would place a limit order not at A, but at the midpoint $\frac{A+B}{2}$. The other sellers then queue below the first seller, just like they do in the one-sided case with $A' = \frac{A+B}{2}$ and $B' = B$.

Without loss of generality, assume that $A = 1$, $B = 0$. Define as before $\varepsilon = r/\lambda$. As before, one can define the state region Ω as the set of states where in equilibrium agents wait a positive expected time.

DEFINITION 3. Let M, μ, f_m, a_m be as in Theorem 2 applied to $A = \frac{1}{2}$ and $B = 0$ (for the bottom half of the limit order book). Define also $g_m = 1 - f_m$ and $b_m = 1 - a_m$, the symmetric values with respect to $1/2$. For $m > M$ define by extension $f_m = 0$ and $g_m = 1$. Let Ω be the collection of points $(0, m)$ and $(m, 0)$ for $m = 0, \ldots, M$.

Theorem 10. *Consider any state (m, n) in the plane. Then a competitive stationary Markov equilibrium of the game is given by the following strategy profile. Because of symmetry, without loss of generality one only needs to describe the strategy of the sellers. Moreover, it is enough to describe the strategy for the bottom seller: if a bottom seller does not follow the appropriate strategy, then a top seller would immediately follow it instead.*

- *If $m > n = 0$, apply the same strategy as in the one-sided case. For example, in state $(m, 0)$ with $0 < m < M$ the bottom seller places a limit order at a_m, and waits. If $m = M$, the bottom seller has a mixed strategy, and exits after Poisson(μ) by placing a market order at B. If $m > M$, the bottom seller immediately places a market order at B and exits. The expected utility of a seller in this state is $f_{m,n} = f_{m-n} = f_m$.*
- *If $m > n > 0$, the bottom seller places immediately a limit order at f_{m-n}, which the top buyer immediately accepts via market order. The expected utility of a seller in this state is $f_{m,n} = f_{m-n}$.*
- *If $n > m > 0$, the bottom seller places immediately a market order at g_{n-m} (but never at a lower level) and exits the game. The expected utility of a seller in this state is $f_{m,n} = g_{n-m} = 1 - f_{n-m}$.*
- *If $m = n > 0$, the bottom seller places immediately a limit order at $f_0 = g_0 = 1/2$, which some buyer immediately accepts via market order. The expected utility of a seller in this state is $f_{m,n} = f_0 = 1/2$.*

PROOF: See Appendix A. □

Note that in each state (m, n), with $m, n > 0$, one has $f_{m,n} = g_{m,n}$. This is because the offers are made at this level, so those who exit get expected utility $f_{m,n}$. Also, those who did not get the offer, move to the state $(m-1, n-1)$ on the same diagonal, where the value functions are by definition the same as in state (m, n).

The equilibrium is not unique, because the offers that agents make to each other can take values in a certain interval (see the proof of the theorem). Nevertheless, the next theorem shows that the equilibrium is essentially unique, in the sense that for all rigid equilibria of this game, the space region degenerates to a subset of the coordinate axes.

Recall that a rigid equilibrium is a competitive stationary Markov equilibrium in which only the traders with the most competitive limit orders use mixed strategies.

Theorem 11. *For any rigid equilibrium in the homogenous case, the state region Ω is a subset of the coordinate axes. As a consequence, the equilibrium limit order book can never contain both buyers and sellers for any positive expected time.*

PROOF: See Appendix A. □

8. Conclusions

This paper presents a tractable model of the dynamics of the limit order book. The shape of the limit order book and its evolution in time are characterized, in several cases in closed form. The traders' optimal choices between submitting a limit order and a market order are also derived.

Having a good model of the limit order book should in principle generate many predictions about the shape of the limit order book, and the evolution of buy and sell prices or of the bid-ask spread. I show that buy and sell orders can cluster away from the bid-ask spread, thus generating a hump-shaped limit-order book. Also, I explain why even in the absence of asymmetric information a market buy order would increase not only the ask price, but also the bid price.

Some of the features of the model point to future avenues of research. One important direction is to study the interaction of agents in this model with a market maker: an agent may decide to have a permanent presence in the market, and take advantage of the liquidity traders in the model—what are then the limits of arbitrage? Another direction is to investigate how the range of prices $[B, A]$ arises endogenously, perhaps as the result of the interaction of traders with different valuations. These limits may change whenever new public information arrives.

Since in this paper trades are usually for one unit, one may wish to incorporate multi-unit trades, and even block trades. Some of this is addressed in the paper, but agents do not have a choice over how many units to trade. A model where agents are strategic about this choice would surely be more complicated, but the payoff is that then one could define a meaningful notion of trading volume and analyze it within the model.

For block trades, perhaps the confines of this type of model would not be enough, and an alternative search model could prove more useful.

Graduate School of Business, University of Chicago, 5807 South Woodlawn Avenue, Chicago, IL 60637, U.S.A.; http://gsb.uchicago.edu/fac/ioanid.rosu.

Appendix A. Proofs of Results

Proof of Proposition 1: One needs to show that there exists a unique solution of the recursive system. First, let us solve the difference equation $2f_m + \varepsilon = f_{m+1} + f_{m-1}$, with initial condition $f_0 = A$. This is a quadratic function

$$f_m = A - bm + \frac{\varepsilon}{2}m^2,$$

with b a constant. The numbers b and M are determined by using $f_M + u\varepsilon = f_{M-1}$ and $f_M = B$. The first equation implies that $f_M - f_{M-1} = \frac{\varepsilon}{2}(2M-1) - b = -u\varepsilon$, so $b = \varepsilon(M - \frac{1}{2} + u)$. Replacing b in the equation $f_M = A - bM + \frac{\varepsilon}{2}M^2 = B$, one obtains

$$A + \varepsilon(\tfrac{1}{2} - u)M - \tfrac{\varepsilon}{2}M^2 = B. \tag{30}$$

The positive solution is

$$M_u = \tfrac{1}{2} - u + \sqrt{(\tfrac{1}{2} - u)^2 + \tfrac{2(A-B)}{\varepsilon}}. \tag{31}$$

Notice that $M_0 = \frac{1}{2} + \sqrt{\frac{1}{4} + \frac{2(A-B)}{\varepsilon}}$ and $M_1 = -\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{2(A-B)}{\varepsilon}}$, so $M_0 - M_1 = 1$. Therefore in the interval $[M_1, M_0)$ there is a unique integer M which corresponds to a number $u \in (0,1]$. This proves both the existence and the uniqueness of a solution of the recursive system.

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 2, 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 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 by $f^0 = f^\infty - r/\lambda$. 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 μ (μ 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. The fact that only equilibria of type 4a and 4b exist is standard. See for example Fudenberg and Tirole (1991, section 4.5.2). □

In the context of Theorem 2, behavior of type 1 appears in states $m > M$; behavior of type 2 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. I show that one gets two more equilibria.

Corollary 13. *In the setup of Proposition 12, assume that the sellers place limit orders as in the context of Theorem 2. Also, the event which 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.*

4 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 13: 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 sellers randomize (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 2: I first prove existence, i.e., that the strategies described by the theorem lead to a competitive stationary Markov equilibrium. Denote by F_m the expected utility of one of the agents in state m. One needs to show that F_m is equal to the previously defined f_m, hence it is the same for all agents in state m.

In each state $m > 1$ there are two types of agents:

- The **bottom** agent, who has the lowest offer in the book, placed at a_m;
- The **top** $m-1$ agents, who are placed above a_m, or at a_m but have arrived after the bottom agent (due to the time priority rule).

Inter-arrival times of Poisson processes are exponentially distributed, so each arrival happens after some random time T, which is an exponential variable with intensity λ. From a state $m = 1, \dots, M-1$, for a top agent the system can go to one of two states:

- $m-1$, if an impatient buyer arrives—after $T_1 \sim \exp(\lambda)$;
- $m+1$, if a patient seller arrives—after $T_2 \sim \exp(\lambda)$.

The first arrival occurs after $T = \min(T_1, T_2) \sim \exp(2\lambda)$, which has expected value $1/2\lambda$. Each event happens with probability $1/2$, so one obtains the formula $F_m = \frac{1}{2}(F_{m-1} + F_{m+1}) - r \cdot \frac{1}{2\lambda}$. Since $\varepsilon = \frac{r}{\lambda}$, the formula for a top seller becomes:

$$2F_m + \varepsilon = F_{m-1} + F_{m+1}.$$

If the agent is the bottom seller, a similar formula is true, except that in state $m-1$ the bottom agent exits the game having sold the asset for a_m. The recursive formula becomes $2F_m + \varepsilon = a_m + F_{m+1}$. Since by definition $a_m = f_{m-1}$, the formula for the bottom seller becomes:

$$2F_m + \varepsilon = f_{m-1} + F_{m+1}.$$

I prove that $F_m = f_m$ by starting with the largest state M. From the state $m = M$, for a top agent the system can go to:

- $M-1$, if an impatient buyer arrives—after T_1;
- M, if a patient seller arrives—after T_2;
- $M-1$, if a current seller places a market order and exits—after T_3.

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. Then, as before, one gets the formula $F_M = F_{M-1} - r \cdot \frac{1}{\lambda+\mu}$. So for a top seller ($u = \frac{\lambda}{\lambda+\mu}$):

$$F_M + u\varepsilon = F_{M-1}.$$

The formula for the bottom seller is $F_M + \frac{\lambda}{\lambda+\mu}\varepsilon = \frac{\lambda}{\lambda+\mu}a_M + \frac{\mu}{\lambda+\mu}B$. Since by definition $a_M = B + \varepsilon$, one can use the above equation to deduce that

$$F_M = B = f_m.$$

Without loss of generality, consider a seller who is the bottom agent in state $m_0 \in \{1, \ldots, M\}$. This implies that the seller can stay in the book only in the states $m_0, \ldots, M$, so one needs to show that $F_m = f_m$ for all $m = m_0, \ldots, M$. If $m_0 = M$, I have already proved that $F_M = B = f_M$. Otherwise, if $m_0 < M$, from the above discussion it follows that F_m satisfies the same system of equations that f_m satisfies. It is easy to see that it is a non-degenerate linear system with $M - m_0 + 1$ equations and unknowns, so $F_m = f_m$ for all $m = m_0, \ldots, M$.

Given that agents in each state m have the same expected utility f_m, it is easy to see that no agent would want to deviate from his strategy. Indeed, in state m if an agent goes below a_m, he would get less than f_m in expectation. And if he is the bottom agent and tries to go above a_m, he will be immediately undercut by some other agents, so he can do no better than f_m. This shows that the equilibrium is subgame perfect. The strategies are clearly Markov and stationary, and since the same argument given above is local, the equilibrium is also competitive. This completes the proof of existence.

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 Appendix B, 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 the 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 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 only depends 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 a positive expected time.

Next, I show 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 $f_M > B$. 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 $f_M = B$.

I now prove by induction that $f_m = B$ for all states $m \geq M$. The above discussion shows that it is true for $m = M$. I just show the first step of the induction: $f_M = B$ implies $f_{M+1} = B$. In state $M+1$, the utility that the other agents get if one agent exits the game is $f^- = f_M = B$. As before, the case $f^0 > f^-$ cannot happen, because then the agents would then wait in state $M + 1$, contradiction. The other possibility is that $f^0 \leq f^-$. But, since $h = B = f^-$, in this case only one equilibrium behavior happens, when some agent exits the game immediately and gets $h = B$. Also, the other agents also have utility $B = f^-$. This is precisely the behavior prescribed by the Theorem in the case when $m > M$. The rest of the induction is proceeds in the same way.

Let us come back to the state $m = M$. From this state, the system goes either 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 13 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 13, so $f_{M-1} = B$, 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 extends the above reasoning to all states $m < M$. This completes the proof of uniqueness. □

Remark 2. It is interesting to see how the utility f_m in the one-sided limit order book depends on B: (recall that $u \in (0, 1]$)

$$\frac{\partial f_m}{\partial B} = \frac{m}{M - \frac{1}{2} + u}. \tag{32}$$

Notice that $\partial f_m / \partial B < 1$ when $m < M$. This formula is important to get intuition about the equilibrium limit order book with both sellers and buyers. Here is a heuristic argument: suppose the system is in an equilibrium with m sellers and n buyers. In that case, the existing sellers in the limit order book can be thought as being in an one-sided equilibrium with B equal to the bid price $g_{m,n}$. One then calculates

$$\frac{\partial f_{m,n}}{\partial n} = \frac{\partial f_{m,n}}{\partial B} \cdot \frac{\partial g_{m,n}}{\partial n} < \frac{\partial g_{m,n}}{\partial n}. \tag{33}$$

This implies that when a new buyer arrives, the sellers are better off by less than the buyers are worse off (recall that $g_{m,n}$ is minus the buyers' utility).

Remark 3. In Section 2, I made the assumption that the impatient traders always submit market orders. Denote by r' the time discount coefficient of the impatient. I now study under what conditions on r' the equilibrium from Theorem 2 still holds when the impatient traders are allowed to behave strategically. The way to extend the equilibrium is to assume that the patient sellers never accept any bid from an impatient buyer. In that case, a buyer is forced to place a market order, and the only question is when. It may be that in certain states it would pay off to defer placing a market order until the arrival of the next state.

The best case scenario for a buyer occurs in state $m = 1$. In that case, the ask is $a_1 = A$, while if the buyer waits, the system goes in one of two states: (i) $m = 0$, where the buyer places a market order at $A = f_0$ (it is assumed that there is always an infinite supply at A from agents outside the model); or (ii) $m = 2$, where the buyer places a market order at $a_2 = f_1$. The gain in expected utility of the buyer from this strategy is (use equation 5 for f_1)

$$\begin{aligned} g - g' &= A - \left(\tfrac{1}{2}(A + f_1) + \tfrac{r'}{2\lambda}\right) = \tfrac{1}{2}(A - f_1) - \tfrac{r'}{2\lambda} \\ &= \tfrac{1}{2}(b - \tfrac{\varepsilon}{2}) - \tfrac{r'}{2\lambda} = \tfrac{\varepsilon}{2}(M - 1 + u) - \tfrac{\varepsilon}{2}\tfrac{r'}{r}. \end{aligned} \tag{34}$$

A necessary and sufficient condition for the gain to be negative is

$$r' \geq r(M + u - 1). \tag{35}$$

Therefore, when equation (35) is true, the equilibrium with strategic impatient buyers is equivalent to the one described by Theorem 2.

What happens if $r' < r(M + u - 1)$? Then the equilibrium can be modified to accommodate the impatient buyer. The first seller only needs to place the first limit order at a level a'_1 which makes the gain of the impatient buyer equal to zero. One calculates

$$a_1' = A - \tfrac{\varepsilon}{2}\big((M-1+u) - \tfrac{r'}{r}\big), \tag{36}$$

and the equilibrium is the same as in Theorem 2, but with $A' = a_1'$ instead of A. One can calculate another formula for the top limit order: $a_1' = (a_1 + a_2)/2 + r'/2\lambda$.

Proof of Proposition 3: First, solve the difference equation $2f_m + \varepsilon = \omega f_{m+1} + \omega f_{m-1}$, with initial condition $f_0 = A$. This has the characteristic equation $\omega x^2 - x + (1-\omega) = 0$, which has two solutions: $x_1 = 1$ and $x_2 = \alpha$, where $\alpha = \frac{1-\omega}{\omega} < 1$. Therefore f_m satisfies

$$f_m = A - C(1-\alpha^m) + \varepsilon\beta m, \tag{37}$$

with $\beta = \frac{1}{2\omega-1}$. The numbers M and C are determined by using $f_M + u\varepsilon = f_{M-1}$ and $f_M = B$. The first equation implies that $f_M - f_{M-1} = C(\alpha^M - \alpha^{M-1}) + \varepsilon\beta = -\varepsilon u$, so

$$C = \frac{\varepsilon(\beta+u)}{\alpha^{M-1} - \alpha^M}. \tag{38}$$

Replacing C in the equation $f_M = A - C(1-\alpha^M) + \varepsilon\beta M = B$, one obtains

$$\frac{A-B}{\varepsilon\beta} = \big(1 + u(2\omega-1)\big)\frac{1-\alpha^M}{\alpha^{M-1}-\alpha^M} - M.$$

Define the right hand side of the above equation by $\phi(M,u)$. One needs to solve $\phi(M,u) = (A-B)/\varepsilon\beta > 0$. Notice that $\phi(M,u)$ is strictly increasing in u. By definition, u belongs to the interval $(u_0, u_1] = (0, \frac{1}{1-\omega}]$. Let us calculate $\phi(M,u)$ when $u = u_1$ (note that $1 + u_1(2\omega - 1) = \frac{\omega}{1-\omega} = \alpha^{-1}$):

$$\phi(M,u_1) = \alpha^{-1}\frac{1-\alpha^M}{\alpha^{M-1}-\alpha^M} - M = \frac{1-\alpha^{M+1}}{\alpha^M - \alpha^{M+1}} - (M+1) = \phi(M+1, u_0). \tag{39}$$

Also, $\phi(1,0) = 0$, and, since $\alpha < 1$, $\lim_{M\to\infty}\phi(M,0) = \infty$. Therefore, the function $\phi(M,u)$, with $M = 0, 1, 2, \ldots$, and $u \in (u_0, u_1]$, sweeps the whole interval $(0,\infty)$ in such a way that every value is attained only once. Define M and u as the unique solution of the equation $\phi(M,u) = \frac{A-B}{\varepsilon\beta}$. This proves both the existence and the uniqueness of a solution to the recursive system.

Given a solution f_m of the recursive system, define by $g_m = f_{m-1} - f_m$ the first difference of f_m. This satisfies the following equations:

$$\begin{cases} \omega g_{m+1} + \varepsilon = (1-\omega) g_m, \\ g_M = u\varepsilon. \end{cases} \tag{40}$$

Hence $g_m = (\omega g_{m+1} + \varepsilon)/(1-\omega)$, and since $g_M = u\varepsilon > 0$, it follows that $g_m > 0$ for $m \le M$. This proves that f_m is strictly decreasing for $m \le M$. Also, $g_{m+1} = \big((1-\omega)g_m - \varepsilon\big)/\omega$, so one can calculate $g_{M+1} = \varepsilon\big((1-\omega)u - 1\big)/\omega \le 0$, since $u \le \frac{1}{1-\omega}$. This shows that f_m is strictly increasing if $m > M$. □

Proof of Proposition 5: The one-sided market with different arrival rates is a Markov system with transition matrix

$$P = \begin{bmatrix} 1-\omega & \omega & 0 & 0 & \cdots & 0 & 0 \\ 1-\omega & 0 & \omega & 0 & \cdots & 0 & 0 \\ 0 & 1-\omega & 0 & \omega & \cdots & 0 & 0 \\ 0 & 0 & 1-\omega & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & \omega \\ 0 & 0 & 0 & 0 & \cdots & 1-\omega & \omega \end{bmatrix}. \tag{41}$$

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. Define

$$\alpha = \frac{1-\omega}{\omega}, \quad \beta = \frac{1}{2\omega - 1}, \quad g = \frac{\omega}{1-\omega} = \alpha^{-1}.$$

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 $(1-\omega)x_{m+1} = \omega x_m$ for all m, hence $x_m = C(\frac{\omega}{1-\omega})^m = Cg^m$. The components x_m must sum to one, so $C = \frac{g-1}{g^{M+1}-1}$. Then

$$x_m = \frac{g^{m+1} - g^m}{g^{M+1} - 1} \approx \alpha^{M-m} - \alpha^{M-m+1}.$$

To calculate the spread $a_m - B$, use the formula $f_{m-1} - f_m = g^{M-m}\varepsilon(\beta + u) - \varepsilon\beta$. From this point forward assume that there is no randomization in state M (so $u = \frac{1}{1-\omega}$).

Then one gets the formula $f_{m-1} - f_m = (g^{M-m+1} - 1)\varepsilon\beta$, which implies

$$\begin{aligned}
a_M - B &= \varepsilon\beta\,(g-1), \\
a_{M-1} - B &= \varepsilon\beta\,\big((g-1) + (g^2-1)\big), \\
a_{M-2} - B &= \varepsilon\beta\,\big((g-1) + (g^2-1) + (g^3-1)\big), \\
&\vdots
\end{aligned}$$

Notice that these spreads appear with stationary probabilities approximately equal to $1-\alpha, \alpha-\alpha^2, \alpha^2-\alpha^3, \ldots$, respectively. It would be interesting to compare this distribution with the empirically observed one (especially at the tails of the distribution).

I give an example of calculating the first moment of the distribution, which is of empirical interest. The exact formula for the average bid-ask spread $\bar{S}$ is

$$\bar{S} = \frac{\varepsilon\beta}{(g-1)(g^{M+1}-1)}\left((M+1)g^{M+3} - (M+3)g^{M+2} + (M+3)g - (M+1)\right).$$

This can be very well approximated by $\bar{S} \approx \varepsilon\beta g M$. Now I give an asymptotic formula for M when ε is small. Recall from (10) that M is obtained by solving $\frac{A-B}{\beta\varepsilon} = \frac{g^{M+1}}{g-1}(1-\alpha^M) - M$. When ε is small, M is large, so g^M is much larger than M. One can drop M from the above formula, and also approximate $1-\alpha^M \approx 1$. One then obtains

$$M \approx \frac{1}{\log g}\log\frac{1}{\varepsilon}.$$

Then the average spread

$$\bar{S} \approx \beta\frac{g}{\log g}\varepsilon\log\frac{1}{\varepsilon},$$

so asymptotically $\bar{S}$ behaves like

$$\bar{S} \sim \varepsilon\log\frac{1}{\varepsilon}.$$

□

Proof of Theorem 8: The proof follows closely the existence part of Theorem 2. Define by $F_{m,n}$ the value function of some seller in state $(m,n) \in \Omega$. One needs to show that $F_{m,n}$ is equal to $f_{m,n}$, hence it is the same for all sellers in state (m,n). Suppose the system is in a state (m,n), which is a boundary point of Ω of Type 5. From this state, the system can go to:

- $(m-1,n)$, if an impatient buyer arrives (after $T_1 \sim \exp(\lambda)$); or if a patient buyer arrives, so the system goes to $(m,n+1)$, and then immediately to $(m-1,n)$ via a fleeting limit order at $f_{m-1,n} = g_{m-1,n}$ (after $T_1' \sim \exp(\lambda)$);
- $(m,n-1)$, if an impatient seller arrives (after $T_2 \sim \exp(\lambda)$); or if a patient seller arrives, so the system goes to $(m+1,n)$, and then immediately to $(m,n-1)$ via a fleeting limit order at $f_{m,n-1} = g_{m,n-1}$ (after $T_2' \sim \exp(\lambda)$);
- $(m-1,n-1)$, if a fleeting limit order is placed at $f_{m,n} = g_{m,n}$ (after $T_3 \sim \exp(\mu_{m,n})$, where $\mu_{m,n} = s_{m,n}\lambda$).

So the system moves to $(m-1,n)$ after $\min(T_1,T_2) \sim \exp(2\lambda)$ (with expected value $1/2\lambda$); or to $(m,n-1)$ in expectation after $1/2\lambda$; or to $(m,n-1)$ in expectation after $1/\mu_{m,n}$. The probabilities by which the system moves to each state are: $2\lambda/(4\lambda+\mu_{m,n})$, $2\lambda/(4\lambda+\mu_{m,n})$, and $s_{m,n}/(4\lambda+\mu_{m,n})$, respectively. Therefore, the utility $F_{m,n}$ of a top seller satisfies $F_{m,n} = \frac{2\lambda}{4\lambda+\mu_{m,n}}(F_{m-1,n}+F_{m,n-1}) + \frac{\mu_{m,n}}{4\lambda+\mu_{m,n}}F_{m-1,n-1}$. Rewrite this as

$$(4+s_{m,n})F_{m,n} = 2F_{m-1,n} + 2F_{m,n-1} + s_{m,n}F_{m-1,n-1}.$$

For the bottom seller, the formula is almost the same, with two exceptions: when an impatient buyer arrives, the bottom seller gets $a_{m,n}$ and exits the game, while the top sellers get $f_{m-1,n}$; and when a fleeting order is placed at $f_{m,n}$, the bottom seller gets $f_{m,n}$, while the top sellers get $f_{m-1,n-1}$. Therefore, the utility $F_{m,n}$ of the bottom seller satisfies $(4+s_{m,n})F_{m,n} = F_{m-1,n} + a_{m,n} + 2F_{m,n-1} + s_{m,n}F_{m,n}$. But by definition $a_{m,n} = f_{m-1,n} + s_{m,n}(f_{m-1,n-1} - f_{m,n})$. So for the bottom seller

$$4F_{m,n} + s_{m,n}f_{m,n} = F_{m-1,n} + f_{m-1,n} + 2F_{m,n-1} + s_{m,n}f_{m-1,n-1}.$$

Notice that the utility of the top sellers and of the bottom seller satisfy formulas which are the same if we replace F by f. So, just as in the proof of Theorem 2, one proceeds by focusing on a particular seller, and show that the seller's utility $F_{m,n}$ satisfies the same system of linear equations that $f_{m,n}$ satisfies. That shows that $F_{m,n} = f_{m,n}$, which is what we wanted to show. The rest of the proof of existence is now straightforward. $\square$

Proof of Theorem 9: I 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

$$\frac{f_{m-1,n} - 2f_{m,n} + f_{m+1,n}}{\delta^2} + \frac{f_{m,n-1} - 2f_{m,n} + f_{m,n+1}}{\delta^2} = 1.$$

But this is the finite difference approximation of the PDE

$$\left(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}\right)(m\delta, n\delta) = 1,$$

which is exactly $\triangle f(x, y) = 1$.

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 $(4 + s_{m,n})f_{m,n} + \varepsilon = 2f_{m-1,n} + 2f_{m,n-1} + s_{m,n}f_{m-1,n-1}$, which after division by δ becomes

$$2\frac{f_{m,n} - f_{m-1,n}}{\delta} + 2\frac{f_{m,n} - f_{m-1,n}}{\delta} + s_{m,n}\frac{f_{m,n} - f_{m-1,n-1}}{\delta} = -\delta.$$

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 $(4 + s_{m,n})f_{m,n} + \varepsilon = f_{m-1,n} + 2f_{m,n-1} + f_{m,n+1} + s_{m,n}f_{m-1,n-1}$, which becomes

$$\frac{f_{m,n} - f_{m-1,n}}{\delta} + \frac{f_{m,n} - f_{m-1,n}}{\delta} + \frac{f_{m,n-1} - 2f_{m,n} + f_{m,n+1}}{\delta^2} \cdot \delta + s_{m,n}\frac{f_{m,n} - f_{m-1,n-1}}{\delta} = -\delta.$$

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. □

Proof of Theorem 10: Using the one-deviation property, assume that in all other states except (m, n) agents behave as mentioned above, and show that for each agent the behavior in state (m, n) is optimal. As mentioned before, without loss of generality one can focus on the behavior of the bottom seller.

Start with the case when $m > n = 0$. When $m > 1$, the bottom seller competes with other sellers for the incoming market orders, and the intuition why this is optimal is the same as in Theorem 2. When $m = 1$, the bottom seller has no reason to have a limit order below $a_1 = f_0 = \frac{1}{2}$, because an incoming buyer is ready to accept it. The limit order cannot be any higher than $\frac{1}{2}$, because in equilibrium an incoming buyer would not accept it.

Suppose $m > n > 0$. This implies that $m \geq 2$. If the bottom seller does not immediately place a limit order at f_{m-n}, then some other seller will. This seller would leave the game together with the top buyer, so the system would move to state $(m-1, n-1)$. In this state, the utility of the remaining sellers is $f_{(m-1)-(n-1)} = f_{m-n}$, so the bottom seller has no incentive to deviate.

Suppose $n \geq m > 0$. If $m \geq 2$, then the bottom seller would not want to deviate, for the same reason as above. If $m = 1$, suppose that $n \geq 2$. As in Proposition 12, define by $g^{\infty}_{1,n}$ (minus) the expected payoff of the buyers if they all wait until the arrival of a new agent. Then the expected utility (taking into account the waiting costs) is

$$g^0_{1,n} = g^{\infty}_{1,n} + \tfrac{\varepsilon}{2} = \tfrac{1}{2}(g_{2,n} + g_{1,n+1}) + \tfrac{\varepsilon}{2} = \tfrac{1}{2}(g_{n-2} + g_n) + \tfrac{\varepsilon}{2} = g_{n-1}.$$

It follows that the seller cannot extract more than $g_{n-1} = b_n$ in that state, because the buyers would not be interested. An alternative would be for the seller to wait until a new arrival. The expected utility for the seller from this strategy would be

$$f^0_{1,n} = \tfrac{1}{2}(f_{2,n} + f_{1,n+1}) - \tfrac{\varepsilon}{2} = \tfrac{1}{2}(g_{n-2} + g_n) - \tfrac{\varepsilon}{2} = g_{n-1}.$$

Again, this strategy would produce nothing better than g_{n-1}, so the seller has no incentive to deviate from placing a market order at $b_n = g_{n-1}$.

Finally, suppose that $m = 1$ and $n = 1$. Then define $f^{\infty}_{1,1}$ the expected payoff of the sellers if they all wait, and by $g^{\infty}_{1,1}$ the same notion for the buyers. One calculates

$$f^{\infty}_{1,1} = \tfrac{1}{2}(f_{1,2} + f_{2,1}) = \tfrac{1}{2}(g_1 + f_1) = \frac{1}{2},$$

since $f_1+g_1=1$ (by definition f and g were chosen symmetric with respect to $\frac{1}{2}$). Then the discounted value to time zero is $f^0_{1,1}=f^\infty_{1,1}-\frac{\varepsilon}{2}$ for the sellers, and $g^0_{1,1}=g^\infty_{1,1}+\frac{\varepsilon}{2}$ for the buyers. Since $f^0_{1,1}<g^0_{1,1}$, it follows that both the buyer and the seller are better off making each other an offer in the interval $[\frac{1}{2}-\frac{\varepsilon}{2},\frac{1}{2}+\frac{\varepsilon}{2}]$. In particular $f_0=g_0=\frac{1}{2}$ belongs to this interval. □

Proof of Theorem 11: Consider a rigid equilibrium of this game. By a proof very similar to that of Theorem 2, one shows that in each state where agents wait for a positive expected time, all sellers have the same utility function (and the same for the buyers).

A state (m,n) in Ω is called *interior* if in that state all agents wait until the arrival of a new agent. A state in Ω is called *boundary* if agents wait for a positive expected time, and at least one trader has a mixed strategy. A state not in Ω is called *fleeting*. Recall that a fleeting limit order is by definition a limit order made quickly by a trader, and immediately accepted by a trader on the other side of the book. A *diagonal* passing through the point (m,n) is defined to be the set of points of the form $\Delta=\{(m+k,n+k)\mid m+k\geq 0, n+k\geq 0, k\in\mathbb{Z}\}$.

Consider a diagonal Δ that intersects Ω. Denote by (M,N) the largest state on Δ which is in Ω. Then the following facts are true:

(a) The state (M,N) is a boundary state, and $f_{M,N}=g_{M,N}$.

(b) All the states $(m,n)\in\Delta$ above (M,N) are fleeting, and $f_{m,n}=g_{m,n}$, which is the same number as $f_{M,N}=g_{M,N}$.

(c) All the states $(m,n)\in\Delta$ below (M,N) are interior, and $f_{m,n}>g_{m,n}$.

If the diagonal Δ does not intersect Ω, then the following fact is true:

(d) All the states $(m,n)\in\Delta$ are fleeting, and $f_{m,n}=g_{m,n}$, which has the same value for all states in Δ.

To prove (a), suppose on the contrary that $f_{M,N}>g_{M,N}$. Since there is no waiting in $(M+1,N+1)$, it follows that there must be a fleeting limit order at h.[22] Using the notation of Proposition 12, one has $f^-=f_{M,N}$, and $g^-=g_{M,N}$. Since $f^->g^-$, either

[22]Technically, it could also happen, for example, that one seller decides to exit by placing a market order at B. But then the same proof can be used as in the case of a fleeting limit order to arrive at a contradiction with $f_{M,N}>g_{M,N}$.

$f^- > h$ or $g^- < h$. Without loss of generality, assume that $f^- > h$. There are two cases: (i) $f^0 > h$: then, as in Case 2 of Proposition 12, agents wait in state $(M+1, N+1)$; (ii) $f^0 \leq h$: this is a game of attrition, so by the rigidity of the equilibrium it follows that agents have to wait in state $(M+1, N+1)$. Both cases lead to a contradiction, so it must be true that $f_{M,N} = g_{M,N}$.

Fact (b) is proved by using essentially the same argument as for (a): since there is no waiting in $(M+k+1, N+k+1)$, then it must be true that $f_{M+k,N+k} = g_{M+k,N+k}$. Moreover, from the proof one notices that the fleeting limit order must be made at $h = f^- = f_{M+k,N+k}$ (otherwise there would be waiting in $(M+k+1, N+k+1)$). But the utility of all sellers in $(M+k+1, N+k+1)$ is h, so $f_{M+k+1,N+k+1} = h = f_{M+k,N+k}$.

To prove (c), I first show that if (m, n) is a boundary state, then $(m-1, n-1)$ is an interior state: From Proposition 12, one knows that mixing is done only when $f^- \geq h \geq f^0$, and $g^- \leq h \leq g^0$. But $f^- = f_{m-1,n-1}$, $g^- = g_{m-1,n-1}$, and $h = f_{m,n} = g_{m,n}$. So one obtains $f_{m-1,n-1} \geq h \geq g_{m-1,n-1}$. But since there is waiting in state (m, n), either $f^{\geq}h$ or $g^{\leq}h$ is a strict inequality. In either case, one gets that $f_{m-1,n-1} > g_{m-1,n-1}$. Since for boundary or fleeting states it has already been shown that $f_{m,n} = g_{m,n}$, it follows that $(m-1, n-1)$ is a interior state. To complete the proof of (c), notice that it was just proved that $(M-1, N-1)$ is interior. Suppose not all the states (m, n) below (M, N) are interior. Then there must be some state (m_0, n_0) which is either boundary or fleeting. But because of (b), in either case all the states above (m_0, n_0) would be fleeting, and one knows that for fleeting states $f = g$. This is in contradiction with $f_{m-1,n-1} > g_{m-1,n-1}$.

Finally, fact (d) is just a consequence of (b), since by definition Ω only contains boundary and interior states, so all the states $(m, n) \in \Delta$ are fleeting.

I now proceed with the proof of the Theorem. By contradiction, suppose that Ω also contains states not on the coordinate axes. Let (m, n) be such a state, which is boundary, and has the extra property that the states $(m+1, n)$ and $(m, n+1)$ are not in Ω. The definition makes sense, because of (a) and the fact that Ω is finite. From (c) it follows that state $(m-1, n-1)$ is interior, therefore

$$f_{m-1,n-1} = \tfrac{1}{2}(f_{m-1,n} + f_{m,n-1}) - \tfrac{\varepsilon}{2}. \tag{42}$$

As in the proof of (c), mixing in a boundary state can be done only when $f^- > h \geq f^0$. In our case, $f^- = f_{m-1,n-1}$, and $f^0 = \frac{1}{2}(f_{m,n+1} + f_{m+1,n}) - \frac{\varepsilon}{2}$. But, because the choice of (m, n), the states $(m+1, n)$ and $(m, n+1)$ are fleeting, so according to (b), $f_{m,n+1} = f_{m-1,n}$, and $f_{m+1,n} = f_{m,n-1}$. Therefore one obtains

$$f_{m-1,n-1} > \tfrac{1}{2}(f_{m-1,n} + f_{m,n-1}) - \tfrac{\varepsilon}{2}. \tag{43}$$

This inequality contradicts formula (42) above, so it follows that Ω must be a subset of the coordinate axes.

□

Appendix B. Continuous Time Game Theory

In this paper the strategic interaction of agents is modeled as a multi-stage game with observed actions in continuous time (for the discrete time version, see Fudenberg and Tirole (1991), ch. 4). The framework I use borrows heavily from the theory of repeated games in continuous time, as developed by Bergin and MacLeod (1993). I extend their framework by allowing stochastic moves by Nature, and entry of new players.

Before I go to the formal definitions, here is a brief discussion about game theory in continuous time. First of all, this theory is not a straightforward extension of discrete time game theory. There are a few conceptual problems, as pointed out by Simon and Stinchcombe (1989), or Bergin and MacLeod (1993). To understand why, suppose one tries to replicate a typical punishment strategy from discrete time repeated games:

> *Continue* to cooperate if the other player has not defected yet; if the other player defected at any point in the past, *immediately* defect and continue to defect forever.

The difficulty to make this strategy precise is two-fold. First, in continuous time there is no first time after t, which makes it difficult to "continue" a certain course. One way to get around this problem is to allow strategies to have inertia. But this creates a second problem, since the other players can take advantage of inertia.[23] One way to allow players to react immediately is to enlarge the concept of strategy to include

[23]One could force the players to all have the same inertia, but then this would be equivalent to forcing the game to take place in discrete time.

sequences of faster and faster responses. The mathematical concept that allows to do that is *completion* with respect to a metric (see below).[24]

A third problem that arises in continuous time is that there is no first time before t. This issue is important when one needs a description of the game right before t. For example, suppose an impatient trader submits a market buy order at t. This market order is supposed to be very fast, and not give time for the existing traders to change their limit orders. One can model this by allowing the market order to be placed "immediately before" t, and the game at t will be played with one less player, namely the trader whose limit order was cleared. Since there is no first time before t, it is not obvious at which price the market order is to be executed. The solution of this problem is to allow only strategies that behave well immediately before any time t. The technical concept, inspired from Simon and Stinchcombe (1989), is of a strategy with a uniformly bounded number of jumps (to be defined below).

As mentioned in Section 2, besides the usual problems with game theory in continuous time, there is an extra problem when dealing with multi-stage games. To wit, suppose a trader submits a market order at t and exits the game. When is then the next stage of the game played? Since in continuous time there is no first time after t, one is compelled to have the next stage-game played also at t. I do this by introducing "layered times," i.e., by allowing multiple games to be played at at the same time.

In order to define a game, one must define the spaces of actions, outcomes, and strategies. The definitions follow closely those of Bergin and MacLeod (1993). I extend their framework in several directions: (i) there is a well-defined description of the game right before any time t; (ii) I allow for entry decisions of new agents; and (iii) I account for the possibility of having more than one game played at the same time. I start with an infinitely countable set of players I. Since they arrive according to independent Poisson processes, with probability one at each point in time there are only finitely many traders. I want to include the case where at some times t the game is played more than once. I do this by taking the product of the time interval $[0, \infty)$ with the set of natural numbers $\mathbb{N}$, to indicate how many time a game has been played at some time t.

[24]Another way to define immediacy is by using infinitesimal numbers, which is the mathematical field of non-standard analysis.

Define

$$\mathcal{T} = [0, \infty) \times \mathbb{N} \tag{44}$$

the set of times at which players can move, counted with multiplicity. Notice that if $\leq$ is the lexicographic order, $(\mathcal{T}, \leq)$ is a totally ordered space. Denote the element $(0, 0) \in \mathcal{T}$ also by 0. Define intervals in $\mathcal{T}$ in the usual way: for example, if $T = (t, n) \in \mathcal{T}$ define $[0, T) = \{T' \in \mathcal{T} \mid 0 \leq T' < T\}$. When there is no danger of confusion, write t instead of (t, n). Also, define a measure on $[0, \infty)$ so that bounded measurable functions are integrable:

$$\mu(\mathrm{d}t) = \mathrm{e}^{-at}\mathrm{d}t. \tag{45}$$

In general, I want the action space for player i to be a compact complete metric space (X_i, d_i). Typically, X_i is a compact subset of $\mathbb{R}^n$ and d_i is the inherited metric. In this paper, the action space for player $i \in I$ can be defined as a subset of $\mathbb{R}^2$:

$$X_i = \big([B, A] \times \{0, 1\}\big) \cup \{\mathbf{out}\}, \tag{46}$$

where **out** is some point in $\mathbb{R}^2$ which does not lie on $[B, A] \times \{0, 1\}$. An action $(x_i, 1) \in X_i$ is interpreted as a limit order at $x_i \in [B, A]$. An action $(x_i, 0) \in X_i$ is interpreted as a market sell (buy) order, in which case x_i is the current bid (ask) price, respectively. The action $x_i = \mathbf{out}$ indicates that either (i) player i has not entered the game yet; or (ii) player i exited the game before. One could also allow agents to exit freely at time t. This will not happen in equilibrium if the utility from exiting is very small, so in order to simplify the description of the game I do not allow free exit. Define also projections on the first and second factor, $\pi_1 : X_i \to [B, A] \cup \{\mathbf{out}\}$ and $\pi_2 : X_i \to \{0, 1\} \cup \{\mathbf{out}\}$, in the obvious way.

I now define outcomes of the game. Let $\mathcal{B}_{X_i}$ and $\mathcal{B}_X$ be the Borel sets of X_i and $X = \prod_{i \in I} X_i$, respectively; and let $\mathcal{B}$ be the Borel sets of $[0, \infty)$. A function $\nu : [0, \infty) \to \mathbb{N}$ is said to have finite support if ν is zero everywhere except on a finite set $M_1 = \{t_1, \ldots, t_K\}$ (its support). One also associates the set $M = \{(t_1, n_1), \ldots, (t_K, n_K)\}$, where all $n_k =$

$\nu(t_k) > 0$. Vice versa, for any such set M one can define a function $\nu_M : [0,\infty) \to \mathbb{N}$ with finite support by sending $t \in [0,\infty)$ to zero if $t \notin M_1$; and to n_k if $t = t_k \in M_1$.

DEFINITION 4. Let X be a space with measure. A function $f : \mathcal{T} \to X$ is called *layered* if there exists a function $\nu : [0,\infty) \to \mathbb{N}$ with finite support such that $\forall t \in [0,\infty)$ and $\forall n, n' > \nu(t)$ one has $f(t,n) = f(t,n')$. If $f : \mathcal{T} \to X$ is layered, associate a function $f^\nu : [0,\infty) \to X$ by $f^\nu(t) = f(t,\nu(t))$. I say that f is a layered measurable function if f^ν is measurable. An *outcome* for player i is a layered Borel measurable function $h_i : \mathcal{T} \to X_i$.

So an outcome is like a regular measurable function $h_i : [0,\infty) \to X_i$, except that at a finite set $\{t_1,\dots,t_K\}$ (the support of ν) it can take several values, up to the integer number $\nu(t_k)$. This corresponds to the idea that at some times t_k the game can be played more than once (in my case, if some agent places a market order).

I call the function ν the *layer* of f. Sometimes I also call the layer of f the associated set $M = \{(t_1,n_1),\dots,(t_K,n_K)\}$, with $n_k = \nu(t_k)$. Also, if f_1 and f_2 are two layered functions with layers ν_1 and ν_2, one can take the combined layer of f_1 and f_2 to be $\nu = \max\{\nu_1,\nu_2\}$. This is useful for situations where one has to compare f_1 and f_2. Consider a layer ν. Then I define: $\mathcal{T}^\nu$, the set of layered times associated with ν; H_i, the space of outcomes for player i; and H_i^ν, the space of outcomes associated with ν:

$$\mathcal{T}^\nu = \{(t,n) \in \mathcal{T} \mid n \le \nu(t)\}, \tag{47}$$

$$H_i = \{h_i : \mathcal{T} \to X_i \mid h_i \text{ layered measurable}\}, \tag{48}$$

$$H_i^\nu = \{h_i : \mathcal{T} \to X_i \mid h_i \text{ layered measurable with layer } \nu\}. \tag{49}$$

This is a metric space with the metric $D_i : H_i^\nu \times H_i^\nu \to \mathbb{R}_+$ given by

$$D_i(h_i,h_i') = \int_{[0,\infty)} d_i\big(h_i^\nu(t), h_i'^\nu(t)\big)\mu(\mathrm{d}t) + \sum_{k=1}^{K}\sum_{n=0}^{\nu(t_k)} d_i\big(h_i(t_k,n), h_i'(t_k,n)\big). \tag{50}$$

Rewrite this as

$$D_i(h_i,h_i') = \int_{\mathcal{T}^\nu} d_i\big(h_i(T), h_i'(T)\big)\, \mu^\nu(\mathrm{d}T). \tag{51}$$

Since the space of measurable functions $f_i : [0, \infty) \to X_i$ is compact and complete, so is H_i^ν. Now, if $\nu \leq \nu'$, there is an inclusion $H_i^\nu \to H_i^{\nu'}$. Also, one knows that for every two layers ν_1 and ν_2 one can take their maximum $\nu = \max\{\nu_1, \nu_2\}$, which satisfies $\nu_1, \nu_2 \leq \nu$. This means that one can regard H_i as the limit of H_i^ν when ν becomes larger and larger. Because of this, H_i is a metric space, but it might not be either complete or compact.

I now define the space H of outcomes of the game. For this, let $H^\nu = \prod_{i \in I} H_i^\nu$ the product space with the metric $D = \prod_{i \in I} \frac{1}{2^i} D_i$. It is a standard exercise in measure theory to see that H^ν is compact and complete. As before, if $\nu \leq \nu'$, there is an inclusion $H^\nu \to H^{\nu'}$. I then define H as the union of all H^ν for all layers ν. This is still a metric space, but it might not be complete or compact. To justify this definition, consider an outcome $h \in H$. Since h belongs to a union of H^ν over all layers ν, there must exist a particular ν so that $h \in H^\nu$ (in which case, I say that ν is the layer of h). This corresponds to the fact that all agents are in the same game, played at the times described by the layer ν.

Also, if $Z \subset \mathcal{T}^\nu$ is layered measurable, and $h_i, h_i' \in H_i^\nu$, define a metric relative to Z by $D_i(h_i, h_i', Z) = \int_Z d_i\big(h_i(T), h_i'(T)\big)\, \mu^\nu(\mathrm{d}T)$. Define also a metric D on H relative to Z in the same way it was done for the product metric above.

Now I define strategies. In discrete time, pure strategies map histories to actions, while mixed strategies map histories to probability densities over actions. For technical reasons it is easier to think of a history as an outcome of the game together with a time t at which history is taken. This way, one can define a strategy as a map from $\{$outcomes $\times$ times$\}$ to $\{$actions$\}$. Formally, a strategy for agent i is a map

$$s_i : H \times \mathcal{T} \to X_i \tag{52}$$

which satisfies the following axioms

A1. The function s_i is layered measurable on $H \times \mathcal{T}$.

A2. For all $h, h' \in H$ and $T \in \mathcal{T}$ such that $D(h, h', [0, T)) = 0$, one has $s_i(h, T) = s_i(h', T)$.

The second axiom ensures that future does not affect current decisions. Rewrite

$$h \sim_T h' \iff D(h, h', [0, T)) = 0.$$

As it was discussed above, these two axioms alone do not ensure that strategies uniquely determine outcomes. For that, one needs some inertia condition. If $t \in [0, \infty)$ and ν is a layer, denote by $t^\nu = (t, \nu(t))$, and $t = (t, 0)$.

A3. The function s_i displays inertia, i.e., for any $h \in H^\nu$ and any $t \in [0, \infty)$, there exists $\varepsilon > 0$ and $x_i \in X_i$ such that

$$D_i\big(s_i(h'), x_i, [t^\nu, t + \varepsilon)\big) = 0$$

for every $h' \in H^\nu$ such that $h \sim_{t^\nu} h'$.

Denote by S_i the set of functions s_i on $H \times \mathcal{T}$ which satisfy A1, A2, A3. Denote by $S = \prod_{i \in I} S_i$. The next theorem shows that a strategy profile $s = (s_i)_i$, i.e., a set of strategies s_i for for each player $i \in I$, uniquely determine an outcome on every subgame. More precisely one has the following result:

PROPOSITION 14. *Let $s \in S$. Then for every $h \in H$ and $T \in \mathcal{T}$, there exists a unique (continuation) outcome $\tilde{h} \in H$ so that $h \sim_T \tilde{h}$ and $D(s(\tilde{h}), \tilde{h}, [T, \infty)) = 0$.*

PROOF: The proof is the same as in Bergin and MacLeod (1993), but one has to make sure that one works in H^ν for some layer ν. □

Given $(h, T) \in H \times \mathcal{T}$ and $s \in S$, denote by $\sigma(s, h, t)$ the outcome which agrees with h on $[0, T)$ and is determined by the strategy s on $[T, \infty)$. Let $s_i, s_i' \in S_i$. I now define a metric on S_i:

$$\rho_i(s_i, s_i') = \sup_{H \times \mathcal{T} \times S_{-i}} D\big(\sigma((s_i, s_{-i}), h, T), \sigma((s_i', s_{-i}), h, T)\big). \tag{53}$$

One also has to introduce an axiom which ensures that for each t the outcome of the game right before t is well defined. One way of doing this is to restrict to strategies s_i that lead to locally constant outcomes with a uniformly bounded number of jumps.

A4. For the strategy s_i there exists M (depending only on s_i) such that for any strategies s_{-i} of the other players, the outcome $\sigma_i\big((s_i, s_{-i}), h, t\big)$ for player i is locally constant and has at most M jumps.

Redefine S_i to include on the strategies that satisfy A4. Now recall that at each $t \in [0, \infty)$ the strategies have inertia for some ε (depending on t). I want inertia to be infinitesimal, because I want to allow for immediate responses. This can be done by completing the space of strategies: Denote by S_i^* the completion of S_i with respect to the metric ρ_i, and by $S^* = \prod_{i \in I} S_i^*$. Completion is done so that the upper bound for the number of jumps is the same for all. More precisely, a point in S_i^* corresponds to a Cauchy sequence $(s_i^n)_n$ of strategies in S_i, and one demands that there exists M so that for each n, s_i^n jumps at most M times, regardless of the other players' strategies. The following result shows that to each strategy in S^* one can associate a unique outcome in H.

Proposition 15. *For every $s \in S^*$ and every $(h, T) \in H \times \mathcal{T}$, there exists a unique h^* such that $\sigma(s^n, h, T) \to h^*$ for any Cauchy sequence $(s^n)_n$ in S converging to s.*

Proof: If $(h, T) \in H \times \mathcal{T}$, there exists a layer ν so that $h \in H^\nu$ and $T \in \mathcal{T}^\nu$. The result then follows easily since H^ν is compact and complete. □

I have just showed that for $s \in S^*$ one can associate a unique outcome of the (whole) game, which I denote by $\sigma^*(s)$. Because completion is done using the same upper bound for the number of jumps, the following result is straightforward. The result allows one to talk about the outcome of a game right before some time t.

Proposition 16. *The outcome $\sigma^*(s)$ associated to a strategy $s \in S^*$ is left-continuous.*

I am almost done in defining the game. The only thing that is left is to describe the payoff for some strategy $s \in S^*$ in a subgame defined by a history $(h, T) \in H \times \mathcal{T}$. Since strategies uniquely define outcomes in every subgame, as long as there exists some payoff $u_i(\sigma^*(s, h, T))$ for each agent i. Define now the equilibrium concept:

Definition 5. A strategy profile $s \in S^*$ is an ε-Nash equilibrium (ε-NE) if for any $h \in H$

$$u_i\big(\sigma^*(s, h, 0)\big) \geq u_i\big(\sigma^*((s_i', s_{-i}), h, 0)\big), \quad \forall i \in I, \quad \forall x_i' \in S_i^*. \tag{54}$$

A strategy profile $s \in S^*$ is an ε-subgame perfect Nash equilibrium (ε-SPE) if for any $(h, T) \in H \times \mathcal{T}$

$$u_i\big(\sigma^*(s, h, T)\big) \geq u_i\big(\sigma^*((s_i', s_{-i}), h, T)\big), \quad \forall i \in I, \quad \forall x_i' \in S_i^*. \tag{55}$$

For $\varepsilon = 0$ in the above inequalities one obtains the concepts of Nash equilibrium (NE) and subgame perfect Nash equilibrium (SPE). One has the following important result.

PROPOSITION 17. *A strategy profile $s \in S^*$ is a subgame perfect equilibrium if and only if for any Cauchy sequence $(s^n)_n$ converging to s, there is a sequence $\varepsilon^n \to 0$ such that s^n is an ε^n-subgame perfect equilibrium.*

PROOF: The proof is essentially the same as in Bergin and MacLeod (1993), but again one has to make sure that one works in H^ν for some layer ν. □

I now discuss mixed strategies. For simplicity of discussion I omit the presence of layers, so one takes $\mathcal{T} = [0, \infty)$. Consistent with this philosophy of locally constant outcomes and inertia strategies, I want to have mixed strategies randomly switch over a small interval. More formally, let X_i be the space of actions for player i, and $[0, \infty]$ the metric space with metric $d(x, y) = |\mathrm{e}^{-x} - \mathrm{e}^{-y}|$. Define a mixed strategy to be a measurable function

$$s_i : H \times \mathcal{T} \to X_i \times X_i \times [0, \infty], \tag{56}$$

where the first component of s_i is the initial action in X_i; the second component is the action to which s_i will switch in the interval of time right after t; the third component is the Poisson intensity of switching. I call this type of strategy "mixed over time," because randomness only comes from the time of switching, while the actions before or after switching are deterministically chosen. One can also allow for mixing over actions, in which case one has to replace X_i by $\Phi(X_i)$, the set of probability densities over X_i, i.e., the set of non-negative integrable functions on X_i with total integral equal to one. Since $\Phi(X_i)$ is compact if X_i is compact, the analysis is essentially the same.

I say that the strategy s_i has inertia in a similar way as before, but one adds the requirement that, after switching, the action to which player i switched will be also held constant for a small period of time. Then one has to modify the description of outcomes, which are now stochastic processes that are built in a very similar way to Poisson processes. The space of strategies is also constructed by taking a completion, in the same way it was done for pure strategies.

To see how a strategy mixed over time works, consider the general case of trading game in this paper, where Nature moves at each time t by bringing new players to the game. One can consider in fact four instances of Nature (four distinct players), each of whom brings one type of trader to the game. The space of actions for each instance of Nature, e.g., the one that brings patient sellers, is the set 2^I of all subsets of I (in principle I allow Nature to add or remove any players from the game). Nature plays the following strategy: if $(h,t) \in H \times \mathcal{T}$ is the history at t, and $J = I_{t_-}$ is the set of players in the game right before t, then $s_N(h,t) = (J, J \cup \{PS\}, \lambda_{PS})$.

I now briefly discuss the notions of equilibrium employed in this paper. The notions of subgame perfect equilibrium and Markov (perfect) equilibrium are simple extensions of the corresponding concepts in discrete time (see Fudenberg and Tirole (1991)). In this paper, I also use the notion of *competitive* Markov equilibrium. This is a Markov perfect equilibrium from which local deviations can be stopped by local punishments, assuming behavior in the rest of the game does not change. In other words, if one truncates the equilibrium strategies by looking at some time interval $(t, t+\delta)$ (truncation is possible because of the Markov condition), the restrictions of the strategies to this time interval remain a Markov equilibrium.

References

[1] Admati, A., and P. Pfleiderer (1988): "A Theory of Intraday Patterns: Volume and Price Variability," *Review of Financial Studies*, 1, 3–40.

[2] Amihud, Y., and H. Mendelson (1980): "Market-Making with Inventory," *Journal of Financial Economics*, 8, 31–53.

[3] Bergin, J., and B. MacLeod (1993): "Continuous Time Repeated Games," *International Economic Review*, 34, 21–37.

[4] Biais, B., P. Hillion, and C. Spatt (1995): "An Empirical Analysis of the Limit Order Book and the Order Flow in the Paris Bourse," *Journal of Finance*, 50, 1655–1689.

[5] Bloomfield, R., M. O'Hara, and G. Saar (2003): "The 'Make' or 'Take' Decision in an Electronic Market: Evidence on the Evolution of Liquidity," Discussion Paper.

[6] Bouchaud, J-P., M. Mezard, and M. Potters (2002): "Statistical Properties of the Stock Order Books: Empirical Results and Models," *Quantitative Finance*, 2, 251–256.

[7] Black, F. (1971): "Towards a Fully Automated Exchange, First Part," *Financial Analysts Journal*, 27, 29–34.

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

[9] Christie, W., and P. Schultz (1994): "Why Do Nasdaq Market Makers Avoid Odd-Eighth Quotes?," *Journal of Finance*, 49, 1813–1840.

[10] Cohen, K., S. Maier, R. Schwartz, and D. Whitcomb (1981): "Transaction Costs, Order Placement Strategy, and Existence of the Bid-Ask Spread," *Journal of Political Economy*, 89, 287–305.

[11] Demsetz, H. (1968): "The Cost of Transacting," *Quarterly Journal of Economics*, 82, 33–53.

[12] Domowitz, I., and A. Wang (1994): "Auctions as Algorithms," *Journal of Economic Dynamics and Control*, 18, 29–60.

[13] Duffie, D., N. Garleanu, and L. Pedersen (2004): "Valuation in Dynamic Bargaining Markets," Discussion Paper, Stanford University, March.

[14] Farmer, D., P. Patelli, and I. Zovko (2003): "The Predictive Power of Zero Intelligence in Financial Markets," Discussion Paper, September.

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

[16] Foucault, T., O. Kadan, and E. Kandel (2004): "Limit Order Book as a Market for Liquidity," *Review of Financial Studies*, forthcoming.

[17] Fudenberg, D., and J. Tirole (1991): *Game Theory*, MIT Press.

[18] Garman, M. (1976): "Market Microstructure," *Journal of Financial Economics*, 3, 257–275.

[19] Glosten, L. (1994): "Is the Electronic Open Limit Order Book Inevitable?," *Journal of Finance*, 49, 1127–1161.

[20] Glosten, L., and P. Milgrom (1985): "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders," *Journal of Financial Economics*, 14, 71–100.

[21] Gladwell, I. and R. Wait (1979), *A survey of numerical methods for partial differential equations*, Clarendon Press, Oxford.

[22] Goettler, R., C. Parlour, and U. Rajan (2004): "Equilibrium in a Dynamic Limit Order Market," *Journal of Finance*, forthcoming.

[23] Grossman, S., and M. Miller (1988): "Liquidity and Market Structure," *Journal of Finance*, 43, 617–633.

[24] Handa, P., and R. Schwartz (1996): "Limit Order Trading," *Journal of Finance*, 51, 1835–1861.

[25] Harris, L. (1998): "Optimal Dynamic Order Submission Strategies in Some Stylized Trading Problems," *Financial Markets, Institutions and Instruments*, Vol. 7, No. 2.

[26] Harris, L., and J. Hasbrouck (1996): "Market versus Limit Orders: the Superdot Evidence on Order Submission Strategy," *Journal of and Quantitative Analysis*, 31, 213–231.

[27] Hasbrouck, J., and G. Sofianos (1993): "The Trades of Market Makers: An Empirical Analysis of NYSE Specialists," *Journal of Finance*, 68, 1565–1593.

[28] Hausman, J., A. Lo, and C. MacKinlay (1992): "An Ordered Probit Analysis of Transaction Stock Prices," *Journal of Financial Economics*, 31, 319–330.

[29] Hollifield, B., R. Miller, P. Sandas, and J. Slive (2003): "Liquidity Supply and Demand in Limit Order Markets," Discussion Paper, Wharton, June.

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

[31] Jain, P. (2002): "Institutional Design and Liquidity on Stock Exchanges around the World," Discussion Paper, Indiana University at Bloomington.

[32] Kumar, P., and D. Seppi (1993): "Limit Orders and Market Orders with Optimizing Traders," Discussion Paper, Carnegie Mellon.

[33] Kyle, Albert (1985): "Continuous Auctions and Insider Trading," *Econometrica*, 53, 1315–1335.

[34] Lo, A., C. MacKinlay, and J. Zhang (2002): "Econometric Models of Limit Order Executions," *Journal of Financial Economics*, 65, 31–71.

[35] O'Hara, M. (1995): *Market Microstructure Theory*, Blackwell.

[36] Parlour, C. (1998): "Price Dynamics in Limit Order Markets," *Review of Financial Studies*, 11, 789–816.

[37] Rock, K. (1996): "The Specialist's Order Book and Price Anomalies," Discussion Paper.

[38] Sandas, P. (2001): "Adverse Selection and Competitive Market Making: Empirical Evidence from a Limit Order Market," *Review of Financial Studies*, 14, 705–734.

[39] Seppi, D. (1997): "Liquidity Provision with Limit Orders and a Strategic Specialist," *Review of Financial Studies*, 10, 103–150.

[40] Simon, L., and M. Stinchcombe (1989): "Extensive Form Games in Continuous Time: Pure Strategies," *Econometrica*, 57, 1171–1214.

[41] Vayanos, D., and T. Wang (2003): "Search and Endogenous Concentration of Liquidity in Asset Markets," Discussion Paper, MIT.