Waiting Costs and Strategic Liquidity Traders in Order-Driven Markets

Juhani Linnainmaa and Ioanid Rosu*

December 1, 2007

Abstract

When liquidity is measured by bid-ask spreads or price impact, markets with more trading activity (such as foreign exchange markets) are typically more liquid than markets with less trading activity (such as municipal bond markets). But showing a causal connection from trading activity to spreads is difficult, due to the endogenous nature of these variables. In the case of Finland's fully electronic limit order market, we show that the weather (measured by sunshine, cloudiness and precipitation) is a valid instrument for trading activity, and that indeed higher trading activity causes lower spreads, both in the time-series and in the cross-section. We further show that weather affects only individual traders, and not institutions.

JEL Classification: C7, etc.

Keywords: Market microstructure, price impact, liquidity trading.

*Graduate School of Business, University of Chicago. juhani.linnainmaa@ChicagoGSB.edu, ioanid.rosu@ChicagoGSB.edu.

Table 1: OLS Panel Regressions of Log-Arrival Rates against Weather and Lagged U.S. Stock Returns

This table uses data from the Helsinki Exchanges in Finland from September 18, 1998 through October 23, 2001 to examine how household investors' and institutional investors' trading activity depends on weather and lagged U.S. stock market returns. A panel OLS regression of log-number of orders against weather and return variables is estimated using daily data for the 44 most actively traded stocks in the market. The weather variable, Sunshine, is the number of hours of sunshine during the day. This variable is demeaned by subtracting off the average number of hours of sunshine in each month (Goetzmann and Zhu 2005) and divided by 100. The U.S. stock return variable, US Ret, is the return on the value-weighted CRSP market index. This variable is lagged by one day and split into positive and negative components; e.g., $(\text{US Ret})^+ = \max(\text{US Ret}, 0)$. The dependent variable is the log-number of orders submitted by household and institutional investors. Separate regression are run for market orders, limit orders, and all orders. Each regression also includes stock-level fixed effects. Each line in this table is a separate regression. The number of stock-day observations in the panel is 29,071.

		Explanatory Variables		
Dependent Variable		Sunshine	$(\text{US Ret})^+$	$(\text{US Ret})^-$
Households	Market Orders	−0.447		
		(0.084)		
		−0.478	4.185	−2.473
		(0.085)	(0.466)	(0.479)
	Limit orders	−0.871		
		(0.129)		
		−0.935	8.391	−4.545
		(0.130)	(0.712)	(0.732)
	All Orders	−0.520		
		(0.089)		
		−0.559	4.931	−2.817
		(0.091)	(0.495)	(0.509)
Institutions	Market Orders	−0.028		
		(0.111)		
		−0.085	6.422	−3.493
		(0.112)	(0.612)	(0.629)
	Limit orders	−0.140		
		(0.121)		
		−0.237	8.480	−4.105
		(0.123)	(0.670)	(0.688)
	All Orders	−0.067		
		(0.115)		
		−0.141	7.353	−3.827
		(0.117)	(0.637)	(0.655)

Table 2: OLS Panel Regressions of Limit-Order Ratio (Competition) against Weather and Lagged U.S. Stock Returns

This table uses data from the Helsinki Exchanges in Finland from September 18, 1998 through October 23, 2001 to examine how household investors' and institutional investors' limit order versus market order activity depends on weather and lagged U.S. stock market returns. A panel OLS regression of the competition variable against weather and return variables is estimated using daily data for the 44 most actively traded stocks in the market. The weather variable, Sunshine, is the number of hours of sunshine during the day. This variable is demeaned by subtracting off the average number of hours of sunshine in each month (Goetzmann and Zhu 2005) and divided by 100. The U.S. stock return variable, US Ret, is the return on the value-weighted CRSP market index. This variable is lagged by one day and split into positive and negative components; e.g., $(\text{US Ret})^+ = \max(\text{US Ret}, 0)$. The dependent variable, competition, is defined as ratio of the number of limit orders to the number of market orders. This ratio is computed separately for buy and sell sides, and the average of the two is used in the regressions. Separate regressions are run for household investors, institutional investors, as well as for all investors. Each regression also includes stock-level fixed effects. Each line in this table is a separate regression. The number of stock-day observations in the panel is 29,071.

	Explanatory Variables		
Dependent Variable	Sunshine	$(\text{US Ret})^+$	$(\text{US Ret})^-$
Household Competition	0.297		
	(0.107)		
	0.302	−2.751	1.694
	(0.109)	(0.596)	(0.613)
Institutional Competition	0.019		
	(0.020)		
	0.025	−0.610	0.168
	(0.020)	(0.109)	(0.112)
All Competition	0.054		
	(0.035)		
	0.081	−1.736	0.808
	(0.035)	(0.193)	(0.198)

Table 3: OLS Panel IV Regressions of Bid-Ask Spread and Price Impact Measures against Household and Institutional Activity

This table uses data from the Helsinki Exchanges in Finland from September 18, 1998 through October 23, 2001 to examine how bid-ask spreads and price impact measures respond to exogenous variation in household and institutional investors' trading activity. A panel OLS IV regression of the bid-ask spread and price impact measure against log-number of orders is estimated using daily data for the 44 most actively traded stocks in the market. The explanatory variable is instrumented with a weather variable (sunshine) and, alternatively, with a weather variable plus lagged U.S. stock market returns. Panel A runs the regressions with the bid-ask spread as the dependent variable. The bid-ask spread is computed as average log-bid-ask spread over the entire trading day, sampled every second. Panel B runs the regressions with a price impact measure as the dependent variable. The price impact measure is defined as follows. First, we compute the following quantity for each price level in the book: the number of shares at the price level divided by the log-distance from the previous price level. (This auxiliary variable is increasing in number of shares and decreasing in the sparseness of the book.) Our price impact measure is the average of these measures for the first 10 price levels in the book. The measure is computed separately for buy and sell sides of the book and then the average of the two measures is taken. A higher number means lower price impact (more liquidity). The dependent variable is the average price impact measure over the entire trading day, sampled every second. Each line in this table is a separate regression. There are 13,732 stock-day observations.

Panel A: Bid-Ask Spread Regressions

	Instruments	
Explanatory Variable	Sunshine	All
Household Activity	−0.687	
	(0.267)	
		−1.069
		(0.167)
Institutional Activity	−17.746	
	(97.222)	
		−0.320
		(0.138)
All Activity	−1.120	
	(0.451)	
		−0.869
		(0.168)

Panel B: Price Impact Regressions

Household Activity	0.915	
	(0.252)	
		0.777
		(0.152)
Institutional Activity	23.617	
	(130.178)	
		−0.057
		(0.135)
All Activity	1.491	
	(0.461)	
		0.472
		(0.155)

Table 4: OLS Panel IV Regressions of Bid-Ask Spread and Price Impact Measures against Household and Institutional Activity and Competition

This table uses data from the Helsinki Exchanges in Finland from September 18, 1998 through October 23, 2001 to examine how bid-ask spreads and price impact measures respond to exogenous variation in household and institutional investors' trading activity and competition (i.e., limit-market order ratio). A panel OLS IV regression of the bid-ask spread and price impact measure against log-number of orders is estimated using daily data for the 44 most actively traded stocks in the market. The explanatory variables are instrumented with a weather variable (sunshine) and lagged U.S. stock market returns. Positive and negative U.S. stock market returns enter as separate variables. Panel A runs the regressions with the bid-ask spread as the dependent variable. The bid-ask spread is computed as average log-bid-ask spread over the entire trading day, sampled every second. Panel B runs the regressions with a price impact measure as the dependent variable. The price impact measure is defined as follows. First, we compute the following quantity for each price level in the book: the number of shares at the price level divided by the log-distance from the previous price level. (This auxiliary variable is increasing in number of shares and decreasing in the sparseness of the book.) Our price impact measure is the average of these measures for the first 10 price levels in the book. The measure is computed separately for buy and sell sides of the book and then the average of the two measures is taken. A higher number means lower price impact (more liquidity). The dependent variable is the average price impact measure over the entire trading day, sampled every second. Each line in this table is a separate regression. There are 13,732 stock-day observations.

Panel A: Bid-Ask Spread Regressions

	Explanatory Variable	
Explanatory Variable	Activity	Competition
Household Investors	−1.567	−1.676
	(0.342)	(0.468)
Institutional Investors	0.205	6.932
	(0.330)	(3.293)
All Investors	−2.276	−5.772
	(0.497)	(1.476)

Panel B: Price Impact Regressions

	Explanatory Variable	
Explanatory Variable	Activity	Competition
Household Investors	1.235	1.541
	(0.301)	(0.412)
Institutional Investors	−0.745	−9.074
	(0.422)	(4.208)
All Investors	1.815	5.509
	(0.439)	(1.305)

References

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

[2] Goetzmann, William, and Ning Zhu (2005): "Rain or Shine: Where is the Weather Effect?" *European Financial Management*, 11, 559–578.

[3] Hong, Harrison, and Jialin Yu (2007): "Gone Fishin': Seasonality in Trading Activity and Asset Prices," Working Paper, October 2007.

[4] Rosu, Ioanid (2007): "A Dynamic Model of the Limit Order Book," Working Paper, University of Chicago.

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 is characterized in closed-form in several cases of interest. The model generates empirically verifiable implications for the shape of the limit-order book and the dynamics of prices and trades. In particular, I show that higher trading activity and higher competition lead to smaller spreads and lower price impact. Moreover, the volatility of the asset should vary in inverse proportion to the square root of trading activity. Also, I show that buy and sell orders can cluster away from the bid-ask spread, thus generating a hump-shaped limit-order book.

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

Date: December 5, 2007.

[1]The author thanks Rob Battalio, Shane Corwin, Thierry Foucault, Drew Fudenberg, Xavier Gabaix, Larry Glosten, Burton Hollifield, Sergei Izmalkov, Eugene Kandel, Leonid Kogan, Jon Lewellen, Andrew Lo, David Musto, Stew Myers, Jun Pan, Christine Parlour, Anna Pavlova, Duane Seppi, Chester Spatt, Richard Stanton, Dimitri Vayanos, and Jiang Wang for helpful comments and suggestions. He is also grateful to participants at the NBER meeting, May 2004; WFA meeting, June 2005; and to seminar audiences at MIT, 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

provides bid and ask quotes. Instead, any sellers (buyers) can place offers (bids) in the 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 waiting costs of the agents in the market, the speed of agents' arrival to the market (trading activity), and the ratio of the numbers of patient to impatient traders (competition).[4]

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*.

[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 ECNs. 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]Trading activity has been shown to explain variation in spreads, starting with Demsetz (1968). Some evidence that reasons other than information may better explain prices can also be found for example

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 arrival rates are a *sufficient statistic* for prices. Here prices change because the arrival of new agents modifies the balance between the various suppliers of liquidity. In particular, the bid-ask spread should be smaller when agents arrive more quickly to the market (more trading activity), or when patient traders arrive faster relatively to the impatient traders (more competition). Put differently, in this framework a market is considered liquid if it is fast and/or competitive. In classical market microstructure a market is liquid if the amount of asymmetric information is small.

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

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").

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 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). In Section 3 I show how to use this formulas to derive robust implications about the average bid-ask spread and price impact, the average the maximum number of traders in the book, and how to use the model to estimate the arrival rates of patient and impatient traders. An interesting implication arising from

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

the model is that the volatility of the asset should vary in inverse proportion to the square of trading activity (the sum of arrival rates of all agents).

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 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, Hillion and Spatt (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 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 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

[6]This effect was observed by Bouchaud, Mezard and Potters (2002, Figure 2). Some evidence of this shape is present in Biais, Hillion and Spatt (1995, p.1664 ff.).

[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.

studied by Parlour (1998), Foucault (1999), Foucault, Kadan and Kandel (2005), Goettler, Parlour and Rajan (2005). 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 (2005). Their setup is similar, except that in their model traders are not allowed to cancel limit orders, and they only care about the bid-ask spread.[8] As a consequence, the model can only provide results about the bid-ask spread. By contrast, my model is able to recover most of their results regarding the spreads and times-to-execution, but also can deal with the whole shape of the limit order book (i.e., the price impact function), and its evolution in time.

Moreover, the main focus of Foucault, Kadan, and Kandel (2005) is on *resilience,* which is defined there as the tendency of the bid-ask spread to revert to small values. But in their paper, as well as in this one, the spread reverts to small values because of the finite number of equilibrium states. In fact, I argue that resilience should depend mainly on how agents' arrival rates vary with the state of the book: if agents arrive fast after the bid-ask spread widened, the market can be considered resilient. Since neither their paper nor mine study state-dependent arrival rates, it is hard to claim that one has truly shed light on resilience.

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

[8]In order to solve the model, Foucault et al. make also two restrictive assumptions regarding the trading process: a buyer must always arrive after a seller, and vice versa; and new orders have to be improving the existing limit orders by at least a tick.

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.[9]

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. Section 4 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 5 describes the equilibrium in the general case with all types of sellers and buyers, and Section 6 concludes.

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 for 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.[10]

[9]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.

[10]In many financial markets, the tick is very small (a penny or even a fraction of a penny). But in the case when the tick size is quite large, a model as in Parlour (1998) could be more appropriate.

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.[11]

Limit orders can be canceled for no cost at any time.[12] 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.[13]

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.

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.

[11]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.

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

[13]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 discount the future in a way proportional to the expected waiting time.[14] 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 similarly 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.[15] It can take only two values: if it is low, the corresponding traders are called *patient*, otherwise they are *impatient*.[16] 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 too much difficulty. From now on I denote by r only the time discount coefficient of the patient agents.

---

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

[15]I prefer to be vague about the exact nature of the waiting costs. Besides the standard time discounting story, one can also think about the opportunity costs of not trading. By keeping their capital tied up in a given position, traders might not be able to pursue potentially more profitable trades. Furthermore, one can interpret waiting costs as uncertainty aversion. Since it is plausible that uncertainty increases with the time horizon, it follows that an uncertainty averse trader loses utility by waiting.

[16]Traditionally, the informed traders are assumed to be the impatient ones (e.g., Glosten (1994). However, Ellul, Holden, Jain and Jennings (2005) provide evidence that orders routed to the automatic execution system of NYSE (as opposed to those routed to the floor auction process) display extreme impatience, yet they appear to be the less information sensitive ones.

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 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.

Setting the game in continuous time nevertheless requires some care.[17] We use the framework developed in Rosu (2006), which allows for multi-stage games and mixed strategies. In general, strategies in continuous time have the property of infinitesimal inertia[18] (as in Bergin and MacLeod (1993)), and have a uniformly bounded number of jumps (as in Simon and Stinchcombe (1989)).

The extension to multi-stage game theory is needed because of market orders: When a market order arrives at time t, an existing limit trader exits the model. Therefore 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 is treated formally in Rosu (2006).

Extra care is also required in defining mixed strategies. Unlike discrete time, in continuous time mixing can be done both over actions, and over time (choosing the time of an action). (See Rosu (2006).) 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.

[17]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.

[18]In continuous-time game theory, agents do not change their strategies in the infinitesimal time interval $[t, t + \mathrm{d}t]$. This is very similar to the setup of continuous-time finance, where agents do not trade during $[t, t + \mathrm{d}t]$.

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 Rosu (2006).

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 14, 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*

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 $(A-B)$ 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.[19] Thus they

[19] See Hollifield, Miller, Sandas and Slive (2004) for an empirical estimation of a model with private valuations.

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 they 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.[20] 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 exogenous, 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 (it is the diagonal effect in Biais, Hillion and Spatt (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 (although Hollifield, Miller, Sandas and Slive (2002) argue that the opposite is true on the Vancouver stock exchange). Moreover,

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

some orders are canceled and never resubmitted. All these empirical facts can be accommodated into the model, and the author has solved some models with these features, but then the number of state variables is larger and the solution is more difficult.

3. Equilibrium: One Side of the Book

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.

Denote the arrival rate of patient sellers by $\lambda_1 = \lambda_{PS}$ and the arrival rate of impatient buyers by $\lambda_2 = \lambda_{IB}$. There are three cases: $\lambda_1 > \lambda_2$, $\lambda_1 = \lambda_2$ and $\lambda_1 < \lambda_2$. In Section 3.3 we will see that the most realistic case corresponds to $\lambda_1 > \lambda_2$, but for now we consider all three of them.

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.[21] 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

[21] 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.

undercut by placing a limit order at $a_2 = A - \delta$ for some very small δ.[22] 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 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. The main intuition is the following: in equilibrium, the existing sellers in the book compete for the incoming market orders from impatient buyers.[23] The sellers have their limit orders

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

[23]The competition among patient traders does not drive expected profits down to zero in our model (see also Biais, Martimort and Rochet (2000)). 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. Harris and Hasbrouck (1996) find a similar result for NYSE's SuperDOT system.

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.[24] 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. This equilibrium is non-competitive because punishment implies that behavior in the rest of the game will be changed (to the competitive equilibrium). Non-competitive equilibria may be important for example in understanding dealer markets.[25] In the present paper I focus on competitive equilibria, since they are the more likely outcome of large, anonymous order-driven markets.

[24]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).

[25]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*

The relevant parameters for are: A; B, which here is the reservation value of the sellers; r, the patience coefficient of the sellers; λ_1, the arrival rate of the patient sellers; and λ_2, the arrival rate of the impatient buyers.

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: otherwise, the expected execution time of the top limit seller would be infinite, hence his utility would be negative infinity; but then he would rather submit a market order at B. 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 $m = 1, \ldots, M-1$, the system can go either to: state $m+1$ if a patient seller arrives—after random time $T_1 \sim \exp(\lambda_1)$; or to state $m-1$ if an impatient buyer arrives—after random time $T_2 \sim \exp(\lambda_2)$. Inter-arrival times of Poisson processes are exponentially distributed, so the arrival of the first of the two states happens at $T = \min(T_1, T_2)$, which is exponential with intensity $\lambda_1 + \lambda_2$ (hence the expected value of T is $\frac{1}{\lambda_1+\lambda_2}$). The first event happens with probability $\frac{\lambda_1}{\lambda_1+\lambda_2}$, while the second event happens with probability $\frac{\lambda_2}{\lambda_1+\lambda_2}$. One obtains the formula

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

Similarly, from the terminal state M, the system can go either to: state M if a patient seller arrives[26]—after random time $T_1 \sim \exp(\lambda_1)$; to state $M-1$, if an impatient buyer arrives—after random time $T_2 \sim \exp(\lambda_2)$; or go to $M-1$, if the seller with the current bottom limit order places a market order at B and exits—after random time $T_3 \sim \exp(\mu)$. Then one obtains the formula

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

Define also: $f_0 = A$.[27] In conclusion, f_m satisfies a system of difference equations: the recursive system.

I now begin the formal discussion of the results. Start with some parameter values A, B, λ_1, λ_2 and r.

[26]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.
[27]This is justified by Theorem 3, in the proof of which one can see that $a_m = f_{m-1}$ for all $1 < m < M$. Since the sole trader at $m = 1$ places an order at $a_1 = A$, one has by extension $A = a_1 = f_0$.

Definition 1. *A sequence f_m, $m = 1, \ldots, M$, is called the* recursive system *associated to A, B, λ_1, λ_2 and r if there is a $\mu \geq 0$ such that the following formulas hold*

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

Proposition 2. *Given A, B, λ_1, λ_2 and r, there exists a unique solution (f_m, M, μ) to the associated recursive system. It satisfies the formulas ($m = 1, \ldots, M$)*

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

$$f_m \; = \; A \; - \; bm \; + \; \tfrac{r}{\lambda_1+\lambda_2}\, m^2, \quad \textit{if } \lambda_1 = \lambda_2. \tag{3}$$

The coefficients $C > 0$ and $b > 0$ satisfy

$$C \; = \; \frac{r}{\lambda_1 - \lambda_2} \, \frac{\frac{\lambda_1+\mu}{\lambda_2+\mu}}{\left(\frac{\lambda_2}{\lambda_1}\right)^{M-1} - \left(\frac{\lambda_2}{\lambda_1}\right)^{M}}, \tag{4}$$

$$b \; = \; \frac{2r}{\lambda_1 + \lambda_2} \left(M - \frac{\mu - \lambda_1}{2(\mu + \lambda_1)} \right). \tag{5}$$

where M is the unique positive integer which for some $\mu \geq 0$ satisfies

$$\frac{A - B}{\frac{r}{\lambda_1-\lambda_2}} \; = \; \frac{\lambda_1 + \mu}{\lambda_2 + \mu} \, \frac{\left(\frac{\lambda_1}{\lambda_2}\right)^M - 1}{\left(\frac{\lambda_1}{\lambda_2}\right) - 1} \; - \; M, \quad \textit{if } \lambda_1 \neq \lambda_2, \tag{6}$$

$$M \; = \; \frac{\mu - \lambda_1}{2(\mu + \lambda_1)} \; + \; \sqrt{\left(\frac{\mu - \lambda_1}{2(\mu + \lambda_1)}\right)^2 + \frac{A - B}{\frac{r}{\lambda_1+\lambda_2}}}, \quad \textit{if } \lambda_1 = \lambda_2. \tag{7}$$

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

Proof. 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 3. *Given $A, B, \lambda_1, \lambda_2, r$, there exists a subgame perfect equilibrium of the game: Let f_m, 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*

$$a_m = f_{m-1}, \quad \text{if } m < M; \tag{8}$$

$$a_M = B + \frac{r}{\lambda_2}. \tag{9}$$

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, and the ask price.[28] *This equilibrium is unique in the class of rigid equilibria, in the sense that any other rigid equilibrium leads to the same evolution of the state variables.*

Proof. See Appendix A. □

Remark 1. 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 (or below), the other sellers can place their limit orders anywhere above a_m. From now on, by an abuse of notation, I consider the equilibria in this class to be the same equilibrium. Moreover, I choose as representative of this class the equilibrium for which there is an infinitesimal cost to canceling a limit order, in the sense that if a trader is indifferent between canceling and not canceling an order, the trader always chooses not to cancel. To define this equilibrium, suppose a new seller arrives when there are already $m - 1$ sellers in the book. Then the strategies require that the new seller place an order at a_m, while the others stay on their previous levels. The outcome of this equilibrium is that, in state m, traders have their offers placed at $a_1, \ldots a_m$, and they never change them.

Also, recall from the assumptions in Section 2 that 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

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

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 3, 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 13 and Corollary 14 in Appendix A.

3.3. *Empirical Implications*

In this section, I derive some empirical implications of the one-sided model. Compared with the the model of Foucault, Kadan and Kandel (2005), where the bid-ask spread is the only state variable, my model also allows one to compute the total number of traders in the book, as well as the price impact function. Moreover, in Foucault, Kadan and Kandel (2005), patient and impatient agents are assumed to arrive at the same rate. Here I allow for different arrival rates, which introduces one more important dimension in which to analyze limit-order markets.

Recall that λ_1, λ_2 are the arrival rates of patient and impatient traders, respectively. We define two more numbers: the *"activity"* parameter λ — the arrival rate of all types of agents; and the *"competition"* parameter a — the ratio of arrival rates of patient to impatient traders:

$$\lambda = \lambda_1 + \lambda_2 = \textbf{activity}, \tag{10}$$

$$c = \frac{\lambda_1}{\lambda_2} = \textbf{competition}. \tag{11}$$

For now I assume that both λ and c are directly observable. Later in this section, I indicate how one can estimate them using the model.

In principle, one should study all three possible cases: $c > 1$, $c = 1$, and $c < 1$. But as will be seen later, given the observed behavior of typical limit order books, the case $c \leq 1$ is unlikely (the average spread would be too wide). So when it comes to empirical predictions, I consider mostly the case $c > 1$.

In the case $c > 1$, I show that the average bid-ask spread and price impact depend on both trading activity and competition in the way one might expect: the higher activity or competition, the smaller the average bid-ask spread and price impact. Moreover, the maximum number of traders in the book increases with trading activity, but surprisingly *decreases* with competition.

The competition measure should in principle also shed more light on the concept of *resilience,* which is the focus of Foucault, Kadan and Kandel (2005). Resilience in their model is the tendency of the bid-ask spread to revert to small values. One can easily show in this model that a higher competition parameter c leads to a more resilient bid-ask spread. But it is hard to argue that resilience is a *robust* implication of either my model or theirs. In both models there exists only a finite number of states, so the limit order book will eventually reach states where the spread is smaller. But this resilience is arguably of second order when compared to the type of resilience that arises in reality.[29]

In what follows, I provide some approximate formulas for the maximum number of limit orders in the book (M), the average bid-ask spread, and the average price impact, as well as their standard deviations.

Proposition 4. *The maximum number of states M satisfies:*

$$M = \frac{\ln\left(\frac{A-B}{r/\lambda}\,\frac{(c-1)^2}{c+1}\right)}{\ln c} + s, \quad \textit{with } s \in (-1, \tfrac{\ln(2)}{\ln(c)}), \qquad \textit{if } c > 1, \tag{12}$$

$$M = \sqrt{\frac{A-B}{r/\lambda}} + s, \quad \textit{with } s \in (-1, 1), \qquad \textit{if } c = 1, \tag{13}$$

$$M = \frac{A-B}{r/\lambda}\,\frac{1-c}{1+c} + s, \quad \textit{with } s \in \left(1, \tfrac{2-c}{1-c}\right), \qquad \textit{if } c < 1. \tag{14}$$

Formula (12) is only true if $\frac{r}{\lambda}$ is sufficiently small, e.g. if $\frac{r}{\lambda} \leq (A-B)\,\frac{c-1}{c+1}\,\frac{\ln(c)}{\ln(4c)}$.

When $c > 1$, the average number of limit traders $\overline{m}$ is approximately equal to M. M is increasing in λ, and decreasing in c (for λ sufficiently large).

Proof. See Appendix A. □

This last result is surprising: one may expect that M should increase with the amount of competition (holding total activity λ constant). This is not the case, because of how the limit orders are placed in the book. When $c > 1$ is relatively large, the spreads between the different limit orders are smaller, but limit orders become more rarefied as

[29] In real limit order markets, arrival rates are not constant but depend on the spread: a wider spread tends to attract limit traders more than a smaller spread, and so the spread reverts to smaller values more quickly than it would by the theoretical channel of eventual convergence. See for example Biais, Hillion and Spatt (1995).

one gets further away from the ask price. On balance, the maximum number of orders M is actually smaller when c is larger.

One may think that by a similar argument the average bid-ask spread $\overline{a_m - B}$ should also be larger when c is large. This would be true if one used the *arithmetic* average across all states. However, here one takes a weighted average with weights given by state probabilities, and these are proportional to c^m. When $c > 1$, smaller spreads are much more likely, therefore the average spread decreases in c.

I now state the formal results. First, introduce some notation: I say that X is asymptotically equal to $f(\varepsilon)$ and write

$$X \approx f(\varepsilon) \quad \text{if} \;\; X = f(\varepsilon) + g(\varepsilon) \;\; \text{and} \;\; \lim_{\varepsilon \to 0} \frac{g(\varepsilon)}{f(\varepsilon)} = 0. \tag{15}$$

(In standard notation one writes $X = f(\varepsilon) + o(f(\varepsilon))$.) In what follows I define

$$\varepsilon = r/\lambda, \tag{16}$$

which is very small if trading activity λ is very high, and/or if the patient agents are very patient (r is small).

Proposition 5. *In the context of Theorem 3, let $S_m = a_m - B$ be the equilibrium bid-ask spread in state m. Then when $\varepsilon = \frac{r}{\lambda}$ is small, the mean spread $\bar{S}$ and the standard* *deviation $\sigma(S)$ can be approximated by*

$$\bar{S} \;\approx\; \varepsilon \ln(1/\varepsilon)\, \tfrac{c(c+1)}{(c-1)\ln(c)}, \quad \sigma(S) \;\approx\; \sqrt{\varepsilon(A-B)}\, \left(\tfrac{(c+1)(c^3+c^2-c)}{(c-1)^3}\right)^{1/2}, \qquad \textit{if } c > 1, \tag{17}$$

$$\bar{S} \;\approx\; \frac{A-B}{3}, \quad \sigma(S) \;\approx\; \frac{2(A-B)}{3}, \qquad \textit{if } c = 1, \tag{18}$$

$$\bar{S} \;\approx\; A-B, \quad \sigma(S) \;\approx\; \varepsilon\, \tfrac{(c+1)\sqrt{1-3c+4c^2-c^4}}{(1-c)^2}, \qquad \textit{if } c < 1. \tag{19}$$

Proof. See Appendix A. □

Notice that when $c = 1$ or $c < 1$, the average spread is too large to be realistic (unless $A - B$ is small, which is less likely if this is estimated in practice by the width of the limit order book). This suggests that indeed in empirical studies one should consider only the case when $c > 1$. In this case, one can check that indeed both $\bar{S}$ and $\sigma(S)$ decrease in λ, and they decrease in c as long as $c < 5$ (which is probably satisfied for most limit order markets).

This proposition is to be compared with Farmer, Patelli and Zovko (2003), who in their cross-sectional empirical analysis of the London Stock Exchange show that with a high R^2 the average bid-ask spread varies proportionally to $\varepsilon^{3/4}$, which is close to our theoretical term $\varepsilon \ln(1/\varepsilon)$.

One more empirical implication about spreads can be derived from Theorem 3:

Proposition 6. *In the context of Theorem 3, there exists a minimum bid-ask spread $S_{\min} = a_M - B$ given by the formula:*

$$S_{\min} \;=\; \frac{r}{\lambda_2} \;=\; \frac{r}{\lambda}\,(a+1). \tag{20}$$

Notice the interesting consequence that, unlike the average bid-ask spread $\bar{S}$, the minimum spread increases in the competition parameter a. This however is not a robust conclusion, since in the two-sided model the minimum bid-ask spread has a much more complicated dependence on the model parameters, and there is no reason why this dependence on a should have the same sign.

One can also identify the volatility of the asset as the standard deviation of the ask price — since the bid price B is constant. Then from Proposition 5, equation (17) we can derive another empirical implication of the model.

Corollary 7. *When $c > 1$, the volatility of the asset $\sigma(a_m)$ varies in inverse proportion to $\sqrt{\lambda}$, the square root of trading activity.*

Of course, in practice the assumption that $A - B$ is constant can only hold for a short interval of time, at most a day. Therefore, this result should only be tested for high-frequency volatility.

Now I consider price impact in this model. For simplicity, I define price impact of one-unit market orders, leaving multi-unit market orders for the next section. According to the model, the price impact of one unit is $a_{m-1} - a_m$, which according to Theorem 3 equals $f_{m-2} - f_{m-1}$ (except for the case when $m = M$). The following result gives an asymptotic formula for the average price impact when $c > 1$.

Proposition 8. *In the context of Theorem 3, define $I_m = -\frac{\mathrm{d}f_{m-1}}{\mathrm{d}m}$ the price impact of a one-unit market order in state m. Then in the case $c > 1$, the mean price impact $\bar{I}$*

and the standard deviation $\sigma(I)$ can be approximated by

$$\bar{I} \;\approx\; \varepsilon \ln(1/\varepsilon)\, \tfrac{c+1}{\ln(c)}, \quad \sigma(I) \;\approx\; \sqrt{\varepsilon(A-B)}\, \sqrt{\tfrac{c+1}{c-1}}. \tag{21}$$

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

One may wonder which empirical implications of the one-sided model are robust, i.e. carry through to the general two-sided model. I conjecture that the asymptotic formulas for the maximum number of traders M, and for the mean and standard deviation of the bid-ask spread and price impact are true, if we replaced asymptotic equality with asymptotic proportionality. The proof of such a statement, however, seems very difficult.

One can also argue that a formula for price impact is more robust than a formula for the bid-ask spread when it comes to generalizing it to the two-sided model. This is because in the one-sided model the price impact is less dependent than the bid-ask spread on the bid price B being constant. (In the two-sided model B can be thought of as moving, as B is replaced by the bid price.)

Finally, we discuss how to estimate the model parameters r, A, B, λ_1 and λ_2. The patience coefficient r cannot really be observed, so it has to be assumed constant for at least a short period of time, and derived from the implications of the model. One can argue also that r should be constant across stocks, as it must depend only on agents' type. The "fundamental band" $[B, A]$ can be proxied simply by looking at the limits of a properly winsorized limit order book.

The difficulty comes from the estimation of λ_1 and λ_2. Simply taking λ_1 as the arrival rate of limit orders, and λ_2 as the arrival rate of market orders is not correct. In the theoretical model λ_1 is the arrival rate of patient agents, and they place limit orders in all states *except* in the state M when the limit order book is full (in state M they behave as if they were impatient and submit market orders). But one notices that the sum $\lambda = \lambda_1 + \lambda_2$ is the same regardless of the equilibrium behavior, so λ can be observed. To estimate $c = \lambda_1/\lambda_2$ from observed data, one cannot simply take the total number of sell limit orders divided by the total number of buy market orders, because that can be showed to theoretically equal one.[30] However, one can take the average competition *conditional* on the various observed spreads. This is the same as the arithmetic average $cM/(M+1)$ which is approximately equal to the theoretical value c.

4. Multi-Unit Market Orders and Price Impact

In the previous section, one saw that with one-unit demands the only limit order price that is fixed in equilibrium is the ask price. It is true that with infinitesimal cancellation costs (as assumed in Remark 1), the other offers above the ask are fixed as well. In that case, a simple calculation shows that the price impact function is convex. However, this conclusion is not robust to the fact that both limit orders and market orders can have more than one unit.

Therefore, in order to determine a more robust distribution of the 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.

[30] In practice, because of cancellations this ratio is not equal to one, and to the extent that cancellations are proportional to λ_1, one may argue that in fact this ratio is a good proxy for competition c.

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* in the book with m limit orders as follows: for each integer $i = 0, \dots, k-1$ define by $Imp(i) = a^{i+1}(m) - a^1(m)$, where a^1 is the ask price, and a^{i+1} is level of the i'th offer above the ask. The price impact represents how much the price moves instantaneously against a trader who submits an i-unit market order. [31]

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.[32] 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.

[31]Notice that there is also a subsequent price impact, since the new ask price does not remain at $a^{i+1}(m)$, but moves to $a^1(m-i)$ which may be different. I will not investigate this subsequent price impact, because it does not seem robust enough to carry through to the general case. Indeed, one can see in Section 5.3 that in the two-sided case the bid price also moves in the same direction as the ask price, and that makes the assumption of B being constant quite restrictive.

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

4.1. *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 $i = 1, \dots, k$. For simplicity, I only study the case when sellers arrive faster than the units demanded by the buyers. This is equivalent to

$$\lambda > \sum_{i=1}^{k} i\,\lambda_i. \tag{22}$$

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:

$$\begin{cases} f_0 = f_{-1} = \cdots = f_{1-k} = A, \\ (\lambda + \sum_{i=1}^{k} \lambda_i) f_m + r = \lambda f_{m+1} + \sum_{i=1}^{k} \lambda_i f_{m-i}, \\ (\sum_{i=1}^{k} \lambda_i + \mu) f_M + r = (\lambda_1 + \mu) f_{M-1} + \sum_{i=2}^{k} \lambda_i f_{M-i}. \\ f_M = B. \end{cases} \tag{23}$$

Similar to the results of Section 3, one obtains the following result:

Proposition 9. *The solution of the recursive system (23) takes the following form*

$$f_m = C_0 + C_1\alpha_1^m + C_2\alpha_2^m + \cdots + C_k\alpha_k^m + \varepsilon\, m, \qquad \textit{where} \tag{24}$$

$$\varepsilon = \frac{r}{\lambda - \sum_{i=1}^k i\lambda_i} > 0 \quad \textit{and} \quad |\alpha_1|, |\alpha_2|, \ldots, |\alpha_k| < 1. \tag{25}$$

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

$$P(X) = \lambda X^{k+1} - (\lambda + \sum_{i=1}^k \lambda_i) X^k + \sum_{i=1}^k \lambda_i X^{k-i}. \tag{26}$$

The constants $C_0, \ldots, C_k$ *are determined uniquely from the first and the last equations of the recursive system.*

The description of the equilibrium is the following:

Theorem 10. *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*

$$a^i(m) = \frac{\lambda_k f_{m-k} + \lambda_{k-1} f_{m-k+1} + \ldots + \lambda_i f_{m-i}}{\lambda_k + \lambda_{k-1} + \ldots + \lambda_i}, \tag{27}$$

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 $m = 2, \dots, M-1$, look at the bottom k levels (or at all m levels if $m < k$), which are $a^1(m), \dots, a^{i_0}(m)$. 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 $m = 2, \dots, M-1$, except for the bottom seller at $a^1(M)$, who exits (by placing a market order at B) after the first arrival in a Poisson process with intensity 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: $m \geq k$. Then $a^k(m) = f_{m-k}$, $a^{k-1}(m) = \frac{\lambda_k f_{m-k} + \lambda_{k-1} f_{m-k+1}}{\lambda_k + \lambda_{k-1}}$, $a^{k-2}(m) = \frac{\lambda_k f_{m-k} + \lambda_{k-1} f_{m-k+1} + \lambda_{k-2} f_{m-k+2}}{\lambda_k + \lambda_{k-1} + \lambda_{k-2}} \dots a^1(m) = \frac{\lambda_k f_{m-k} + \lambda_{k-1} f_{m-k+1} + \cdots + \lambda_1 f_{m-1}}{\lambda_k + \lambda_{k-1} + \cdots + \lambda_1}$. Since it does not matter what happens above $a^k(m)$, one can also choose by convention: $a^{k+1}(m) = f_{m-k-1}$, ..., $a^{m-1}(m) = f_1$, $a^m(m) = f_0$.
- CASE 2: $m < k$. Then $a^m(m) = \frac{\sum_{i=m}^k \lambda_i A}{\sum_{i=m}^k \lambda_i} = A$, $a^{m-1}(m) = \frac{\sum_{i=m}^k \lambda_i A + \lambda_{m-1} f_1}{\sum_{i=m}^k \lambda_i + \lambda_{m-1}} \dots$ $a^1(m) = \frac{\sum_{i=m}^k \lambda_i A + \lambda_{m-1} f_1 + \cdots + \lambda_1 f_{m-1}}{\sum_{i=m}^k \lambda_i + \lambda_{m-1} + \cdots + \lambda_1}$.

4.2. *Price Impact of Transactions: Numeric Results*

As pointed out in the introduction to this section, in the case of multi-unit market orders it may be optimal for agents to cluster away from the ask. What determines the

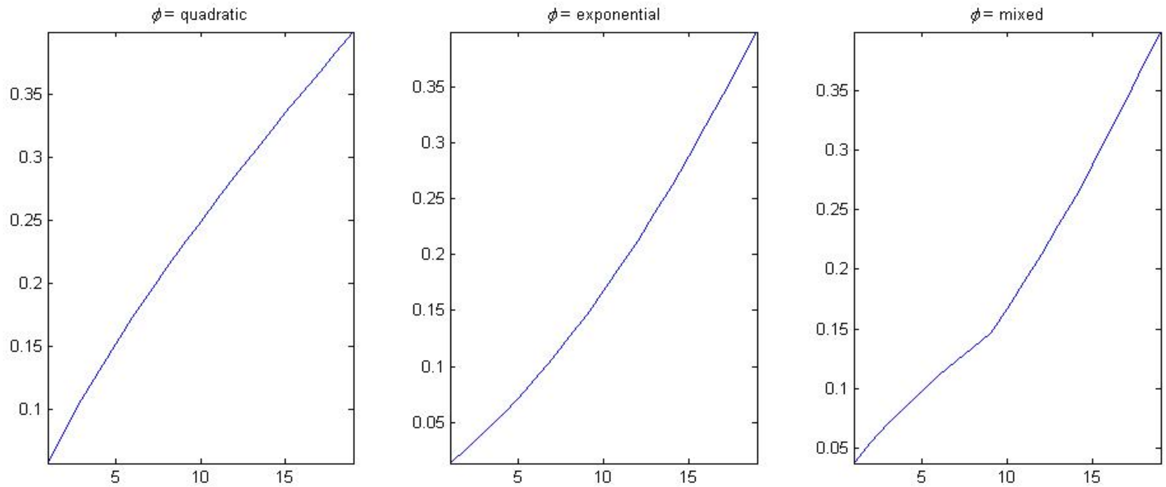

Figure 1. The instantaneous price impact function $Imp(i)$ plotted against i when the decrease ϕ of market order size probabilities is: (i) ϕ_1=quadratic; (ii) ϕ_2=exponential; and (iii) ϕ_3=mixed: first quadratic and then exponential. $Imp(i)$ is the difference between the level of the i'th limit sell order above the ask, and the ask price. The 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 limit order book is in state $m = 30$ (the number of limit orders in the book). The graphs are drawn only up to $k = 20$, the maximum number of units that market buy orders can trade with positive probability. The shape of the equilibrium price impact function is: (i) concave, (ii) convex, and (iii) first concave then convex.

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.[33]

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:

$$Imp(i, m) = a^{i+1}(m) - a^{1}(m), \quad \text{as a function of } i \text{ (and } m). \tag{28}$$

[33] 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.

To calculate the price impact function in a concrete example, one can apply the theoretical results of this section in the following way: Equation (23) shows that f_m satisfies a recursive formula with initial conditions $f_0 = f_{-1} = \cdots = f_{1-k} = A$. This fixes k coefficients of f_m in equation (24). To fix the last coefficient, one should use the equations in (23). For simplicity, I employ an alternative 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.[34] Then one determines B by $B = f_M$. In the analysis that follows I use the parameter δ instead of B. Consider the following values for the numerical application:

$$A = 1, r = 0.001, \delta = 0.04.$$

When one analyzes the shape of the price impact function, it turns out that a crucial factor is how fast the arrival rates λ_i decrease in i. For example, consider a function $\phi(i)$ which is decreasing in i, and some small value $\lambda_0 > 0$ such that

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

[34] As in Propositions 2, one verifies that f_m decreases for $m < M$ and increases for $m > M$.

Then set the arrival rate of the patient sellers $\lambda = \sum_{i=1}^{k} i\lambda_i$.[35] In the numerical application, I choose $\lambda_0 = 10^{-5}$, which is very small, to illustrate the fact that the shape of the price impact function depends on the *relative* magnitude of the arrival probabilities.

In Figure 1, I compare the graphs of the price impact function $Imp(i, m)$ for the the following three functions:

(i) $\phi_1(i) = 1/i(i+1)$;

(ii) $\phi_2(i) = 10/10^i$;

(iii) $\phi_3(i) = \begin{cases} 1/i(i+1) & \text{if} \quad i = 1, \ldots, 10, \\ C/10^i & \text{if} \quad i = 10, \ldots, 20. \end{cases}$

The constant C is such that ϕ_3 is well defined for $i = 10$.

The limit order book is in the state where there are $m = 30$ limit orders. The graphs are only displayed up to $k = 20$, since market orders of size $i > k$ appear with zero probability (hence it does not matter what sellers do above the k'th level from the ask).

Notice that for ϕ_1 the price impact function $Imp(i)$ is concave, while for ϕ_2 it is convex. For ϕ_3, the price impact function is concave for $i < 10$, and convex for $i >= 10$. The intuition for the latter result is the following: sellers up to level 10 expect that their limit orders will be cleared by a market order with a probability that 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 level 10, the probability of a large market order decreases fast, so the sellers prefer to cluster near level 10. The price impact function above that level becomes convex.

---

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

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.[36] 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.[37]

5. 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 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

[36] 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.

[37] Huberman and Stanzl (2000) argue that a non-linear price impact function leads to arbitrage. Empirical studies typically indicate a concave price function, e.g., Hasbrouck (1991), Hausman, Lo and MacKinlay (1992), Keim and Madhavan (1996), Knez and Ready (1996). If one eliminates the influence of large or block trades, the price impact function is almost linear, e.g., Breen, Hodrick and Korajczyk (2002), and Sadka (2003).

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.

5.1. *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

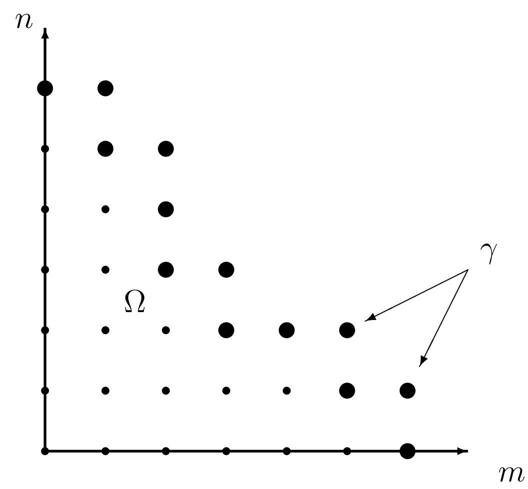

FIGURE 2. The state region Ω and the boundary γ.

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.

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.[38]

[38]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, these types of points will be ignored.

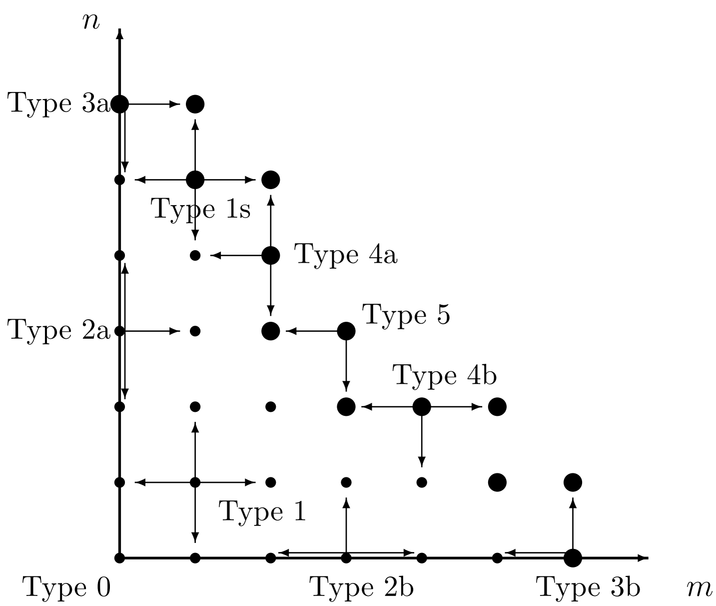

FIGURE 3. Types of points in the state region Ω.

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;
- $(m-1, n-1)$, 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 parameters are A, B, r, and λ. Define

$$\varepsilon = r/\lambda. \tag{29}$$

As in Definition 1, we define $f_{m,n}$, $g_{m,n}$, and $a_{m,n}$, $b_{m,n}$ in Definition 15 in the Appendix. The main result of this section is that, given an appropriate solution of the recursive system, there exists an equilibrium of the game.

Theorem 11. *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 15. 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.

I was not able to prove this result, although numerically it verified in all the cases that were studied.

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 15 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 $f(x, y) = f_{m,n}, g(x, y) = g_{m,n}$. Then one obtains the following asymptotic result:

Theorem 12. *Any solution of the recursive system in Definition 15 converges when* $\varepsilon = r/\lambda$ *is small to the solution of the following system of partial differential equations* *with a free boundary γ:*

$$\tag{30} \begin{cases} \triangle f = 1, \\ f(0,y) = A, \\ \frac{\partial f}{\partial y}(x,0) = 0, \\ \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} = 0 \ \text{at } \gamma; \end{cases} \qquad \begin{cases} \triangle g = -1, \\ g(x,0) = B, \\ \frac{\partial g}{\partial x}(0,y) = 0, \\ \frac{\partial g}{\partial x} + \frac{\partial g}{\partial y} = 0 \ \text{at } \gamma; \end{cases}$$

where the free boundary γ is determined by the condition

$$\tag{31} f = g \ \text{at } \gamma.$$

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).

5.2. *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 12. 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).

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.

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

An implication of the equilibrium in the general case is that a market sell order leads to a decrease in both the bid and the ask, but the decrease in the bid is larger. The intuition for this is the following: a market sell order clears the top limit buy orders, which makes the buyers immediately better off, because there is less competition among

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

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

them. At the same time, the sellers are worse off, because their reservation value (which is the bid price) decreased. But the influence of this reservation value manifests itself directly only later, when the limit order book is full. Therefore, the decrease of the bid is bigger than the decrease of the ask, so the bid-ask spread widens.

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

This phenomenon was noticed empirically by Biais, Hillion and Spatt (1995), in their analysis of the order flow in the Paris Bourse (now Euronext). They attribute it to information: a decrease in the bid could be mechanical, but the decrease in the ask must be due to the information arising from the downward shift in the expected fundamental value. 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. My model is not the only one where this phenomenon occurs in the absence of information. Another model is given by Parlour (1998). The contribution of the present model is to show how waiting costs and competition can also lead to such a phenomenon, even if there is no asymmetric information.

6. 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 higher trading activity and higher competition among limit traders generate smaller spreads and lower price impact. If one interprets the volatility of the asset as the volatility of the ask price, it must vary in inverse proportion to the square root of trading activity. Also, because of the small probability of multi-unit market orders, buy and sell limit orders can cluster away from the bid-ask spread, thus generating a hump-shaped limit-order book.

Some of the limitations 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 2: One needs to show that there exists a unique solution of the recursive system. Let $\lambda = \lambda_1 + \lambda_2$, $a = \frac{\lambda_1}{\lambda_2}$, and $\varepsilon = \frac{r}{\lambda}$.

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

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

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

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

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

Finally, I prove the last statement of the Proposition in the case $c = 1$, for the other cases the proof being similar. 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 3, 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 13. *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 $f^0 \le f^-$, 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 13: 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 $f^- \le h$, 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 3, 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 14. *In the setup of Proposition 13, assume that the sellers place limit orders as in the context of Theorem 3. 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 13. Then, if $h \in [f^0, f^-]$ and $f^0 \leq f^-$, besides the equilibria in Proposition 13, 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 14: 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 3: 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_i. From a state $m = 1, \ldots, M-1$, for a top agent the system can go to one of two states:

- $m-1$, if an impatient buyer arrives—after $T_1 \sim \exp(\lambda_1)$;

- $m+1$, if a patient seller arrives—after $T_2 \sim \exp(\lambda_2)$.

The first arrival occurs after $T = \min(T_1, T_2) \sim \exp(\lambda_1 + \lambda_2)$, which has expected value $\frac{1}{\lambda_1+\lambda_2}$. One obtains the formula $F_m = \frac{\lambda_1}{\lambda_1+\lambda_2} F_{m+1} + \frac{\lambda_2}{\lambda_1+\lambda_2} F_{m-1} - r \cdot \frac{1}{\lambda_1+\lambda_2}$. The formula for a top seller becomes: $(\lambda_1 + \lambda_2) F_m + r \ = \ \lambda_1 F_{m+1} + \lambda_2 F_{m-1}$. 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 $(\lambda_1 + \lambda_2) F_m + r = \lambda_1 F_{m+1} + \lambda_2 a_m$. Since by definition $a_m = f_{m-1}$, the formula for the bottom seller becomes: $(\lambda_1 + \lambda_2) F_m + r \ = \ \lambda_1 F_{m+1} + \lambda_2 f_{m-1}$.

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 $T_1 \sim \exp(\lambda_1)$;
- M, if a patient seller arrives—after $T_2 \sim \exp(\lambda_2)$;
- $M-1$, if a current seller places a market order and exits—after $T_3 \sim \exp(\mu)$.

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, as before, one gets the formula $F_M = F_{M-1} - r \cdot \frac{1}{\lambda_2+\mu}$. So for a top seller $F_M + \frac{r}{\lambda_2+\mu} = F_{M-1}$. The formula for the bottom seller is $F_M + \frac{r}{\lambda_2+\mu} = \frac{\lambda_2}{\lambda_2+\mu} a_M + \frac{\mu}{\lambda_2+\mu} B$. Since by definition $a_M = B + \frac{r}{\lambda_2+\mu}$, 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 Rosu (2006), strategies must satisfy Property A4 (they must have finitely many jumps), which implies that the state variables are left-continuous. In particular, it makes sense to talk about the value of the state variables (m, a_m) right before some time t.

Consider a restriction of the strategies to some time interval $(t, t + \delta)$, such that no strategy has a jump during that interval. This restriction can be made because of the Markov condition: history is reduced to the limit of outcomes at a single point. On this interval, all agents have the same utility: Otherwise, suppose 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 13, 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 $f^0 \le B$. 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 $f^0 \le B$, 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 $m \ge M$. 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 14 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 14, 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. $\square$

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

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

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

When $c > 1$, one can compute the average number of limit of traders $\overline{m}$ using the methods from the proof of Proposition 5. The state m appears with probability proportional to c^m, so $\overline{m} = \frac{1\times 0 + c\times 1 + c^2 \times 2 + \cdots + c^M \times M}{1+c+c^2+\cdots+c^M} = \frac{c\left(Mc^{M+1} - (M+1)c^M + 1\right)}{(c-1)(c^{M+1}-1)} = M -$ o(M), where $o(M)/M \to 0$ when $M \to \infty$.

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

Proof of Proposition 5: For simplicity, assume that the trader at the ask does not have a mixed strategy, i.e. that $\mu = 0$. The one-sided market with different arrival rates is a Markov system with transition matrix

$$P = \begin{bmatrix} \frac{1}{c+1} & \frac{c}{c+1} & 0 & 0 & \cdots & 0 & 0 \\ \frac{1}{c+1} & 0 & \frac{c}{c+1} & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{c+1} & 0 & \frac{c}{c+1} & \cdots & 0 & 0 \\ 0 & 0 & \frac{1}{c+1} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & \frac{c}{c+1} \\ 0 & 0 & 0 & 0 & \cdots & \frac{1}{c+1} & \frac{c}{c+1} \end{bmatrix}. \tag{32}$$

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

To calculate the average bid-ask spread $a_m - B$, notice that the spread is $A - B$ in state $m = 0$, $f_{m-1} - B$ in state $m = 1, \ldots, M-1$, and $\frac{r}{\lambda_2}$ in state $m = M$. To compute then the mean and standard deviation of the spread, one needs to perform a tedious calculation that I will omit. To give just one example, when $c > 1$, the average spread can be computed to be precisely $\overline{a_m - B} = \frac{r/\lambda\,(c+1)}{(c-1)^2(c^{M+1}-1)}\big((M-1)c^{M+3} - (M-2)c^{M+2} - 3c^{M+1} + c^2 + (M+1)c - M\big) \approx \frac{r/\lambda\, c(c+1)}{c-1} M$. But from Proposition 4, $M \approx \frac{\ln\left(\frac{1}{r/\lambda}\right) + \ln\left((A-B)\frac{(c-1)^2}{c+1}\right)}{\ln(c)} \approx \frac{\ln\left(\frac{1}{r/\lambda}\right)}{\ln(c)}$. Since $\varepsilon = \frac{r}{\lambda}$, one gets $\overline{a_m - B} \approx \frac{c(c+1)\varepsilon \ln\left(\frac{1}{\varepsilon}\right)}{(c-1)\ln(c)}$. □

Proof of Proposition 8: Similar to the proof of Proposition 5, one needs to consider the average price impact over states m which appear with probabilities proportional to c^m. One computes $I_m = f_{m-2} - f_{m-1} = -\frac{r}{\lambda_1 - \lambda_2} + C\left(\left(\frac{\lambda_2}{\lambda_1}\right)^{m-2} - \left(\frac{\lambda_2}{\lambda_1}\right)^{m-1}\right) = \frac{r}{\lambda_1-\lambda_2}(c^{M+2-m} - 1)$, and after some tedious calculations one gets the desired formulas. □

Definition 15. *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.

Proof of Theorem 11: The proof follows closely the existence part of Theorem 3. 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 3, 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. □

Proof of Theorem 12: 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. □

References

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

[2] 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.
[3] Biais, B., D. Martimort, and J-C. Rochet (2000): "Competing Mechanisms in a Common Value Environment," *Econometrica*, 68, 799–837.
[4] Bloomfield, R., M. O'Hara, and G. Saar (2005): "The 'Make or Take' Decision in an Electronic Market: Evidence on the Evolution of Liquidity," *Journal of Financial Economics*, 75, 165–199.
[5] 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.
[6] Breen, D., L. Hodrick, and R. Korajczyk (2002): "Predicting Equity Liquidity," *Management Science*, 48, 470–483.
[7] Chakravarty, S., and C. Holden (1995): "An Integrated Model of Market and Limit Orders," *Journal of Financial Intermediation*, 4, 213–241.
[8] Christie, W., and P. Schultz (1994): "Why Do Nasdaq Market Makers Avoid Odd-Eighth Quotes?," *Journal of Finance*, 49, 1813–1840.
[9] 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.
[10] Demsetz, H. (1968): "The Cost of Transacting," *Quarterly Journal of Economics*, 82, 33–53.
[11] Duffie, D., N. Garleanu, and L. Pedersen (2004): "Valuation in Dynamic Bargaining Markets," Discussion Paper, Stanford University, November.
[12] Dufour, A., and R. Engle (2000): "Time and the Price Impact of a Trade," *Journal of Finance*, 55, 2467–2498.
[13] Ellul, A., C. Holden, P. Jain, and R. Jennings (2005), "Order Dynamics: Recent Evidence from the NYSE," working paper, May.
[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 (2005): "Limit Order Book as a Market for Liquidity," *Review of Financial Studies*, 18, 1171–1217.
[17] Fudenberg, D., and J. Tirole (1991): *Game Theory*, MIT Press.
[18] Glosten, L. (1994): "Is the Electronic Open Limit Order Book Inevitable?," *Journal of Finance*, 49, 1127–1161.
[19] 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.
[20] Gladwell, I. and R. Wait (1979), *A survey of numerical methods for partial differential equations*, Clarendon Press, Oxford.
[21] Goettler, R., C. Parlour, and U. Rajan (2005): "Equilibrium in a Dynamic Limit Order Market," *Journal of Finance*, 60, 2149–2192.
[22] Handa, P., and R. Schwartz (1996): "Limit Order Trading," *Journal of Finance*, 51, 1835–1861.
[23] 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.
[24] Hasbrouck, J., and G. Sofianos (1993): "The Trades of Market Makers: An Empirical Analysis of NYSE Specialists," *Journal of Finance*, 68, 1565–1593.
[25] Hasbrouck (1991): "Measuring the Information Content of Stock Trades," *Journal of Finance*, 66, 179–207.
[26] Hausman, J., A. Lo, and C. MacKinlay (1992): "An Ordered Probit Analysis of Transaction Stock Prices," *Journal of Financial Economics*, 31, 319–330.
[27] Hollifield, B., R. Miller, P. Sandas, and J. Slive (2002): "Liquidity Supply and Demand in Limit Order Markets," CEPR Discussion Paper, December.

[28] Hollifield, B., R. Miller, P. Sandas, and J. Slive (2004): "Empirical Analysis of Limit Order Markets," *Review of Economic Studies*, 71, 1027–1063.
[29] Huang, R., and H. Stoll (1997): "The Components of the Bid-Ask Spread: A General Approach," *Review of Financial Studies*, 10, 995–1034.
[30] Huberman, G., and W. Stanzl (2000): "Arbitrage-Free Price-Update and Price-Impact Functions," Yale University Discussion Paper, October.
[31] Jain, P. (2002): "Institutional Design and Liquidity on Stock Exchanges around the World," Discussion Paper, Indiana University at Bloomington.
[32] Keim, D., and A. Madhavan (1996): "The Upstairs Market for Large-Block Transactions: Analysis and Measurement of Price Effects," *Review of Financial Studies*, 9, 1–36.
[33] Knez, P., and M. Ready (1996): "Estimating the Profits from Trading Strategies," *Review of Financial Studies*, 9, 1121–1163.
[34] Kyle, Albert (1985): "Continuous Auctions and Insider Trading," *Econometrica*, 53, 1315–1335.
[35] Lo, A., C. MacKinlay, and J. Zhang (2002): "Econometric Models of Limit Order Executions," *Journal of Financial Economics*, 65, 31–71.
[36] O'Hara, M. (1995): *Market Microstructure Theory*, Blackwell.
[37] Parlour, C. (1998): "Price Dynamics in Limit Order Markets," *Review of Financial Studies*, 11, 789–816.
[38] Rock, K. (1996): "The Specialist's Order Book and Price Anomalies," Discussion Paper.
[39] Rosu, I. (2006): "Multi-Stage Game Theory in Continuous Time," Working Paper, University of Chicago.
[40] Sadka, R. (2003): "Momentum, Liquidity Risk, and Limits to Arbitrage," job market paper, January.
[41] Sandas, P. (2001): "Adverse Selection and Competitive Market Making: Empirical Evidence from a Limit Order Market," *Review of Financial Studies*, 14, 705–734.
[42] Seppi, D. (1997): "Liquidity Provision with Limit Orders and a Strategic Specialist," *Review of Financial Studies*, 10, 103–150.
[43] Simon, L., and M. Stinchcombe (1989): "Extensive Form Games in Continuous Time: Pure Strategies," *Econometrica*, 57, 1171–1214.
[44] Vayanos, D., and T. Wang (2003): "Search and Endogenous Concentration of Liquidity in Asset Markets," Discussion Paper, MIT.