Liquidity shocks and order book dynamics*

Bruno Biais[†] and Pierre-Olivier Weill[‡]

January 25, 2009

Abstract

We propose a dynamic equilibrium model of limit order trading, based on the premise that investors submit limit orders because they can't monitor the market continuously. We study how our theoretical limit order market reacts to a transient liquidity shock, when a significant fraction of investors loose their willingness and ability to hold the asset. We characterize analytically the equilibrium dynamics of market prices, bid-ask spreads, order submissions and cancelations strategies, as well as the volume and limit order book depth they generate. A comparative static exercise shows that, when investors' ability to monitor the market improves, the ratio of messages (order submission and cancelations) to volume increases, consistent with recent evidence on the impact of computerization and algorithmic trading.

*Many thanks, for helpful discussions and suggestions, to Darrell Duffie, Emmanuel Farhi, Hugo Hopenhayn, John Moore, Henri Pages, Jean Charles Rochet, Jean Tirole, Dimitri Vayanos and Glen Weyl; and seminar participants at the Federal Reserve Bank of Minneapolis, UCLA, Toulouse, LSE, LBS, the Banque de France Conference on Macroeconomics and Liquidity, the University of Minnesota Carlson School of Business, UCLA Anderson, Yale University, and Harvard University.

[†]Toulouse School of Economics.

[‡]University of California Los Angeles and NBER.

Investors are not continuously trading. Final investors have full time jobs and other matters to attend. Professional traders too have other tasks to fulfill: they must participate in meetings, interact with customers and comply with reporting requirements. Furthermore, in order to make efficient financial decision, they need to collect and process information regarding assets supply and demand, as well as the fundamentals of underlying cash flows. All these activities take time away from trading.

That all investors are not permanently trading reduces liquidity (Demsetz (1968), Garbade and Silber (1979)). But investors can use the order book, to leave limit orders in the market. In the words of Harris (2003): "Limit orders represent absent traders [enabling them] to participate in the markets while they attend to business elsewhere."

The goal of this paper is to analyze the equilibrium dynamics of the order book in this context. We focus on market dynamics following liquidity shocks. Liquidity shocks arise when a significant fraction of the investors' population is affected by a change in its willingness and ability to hold the asset, as in Grossman and Miller (1988). This can be due to changes in the characteristics of the assets. For example, many institutions are required to sell bonds when they loose their investment grade status, or stocks when they are de-listed from exchanges or indices (see Greenwood, 2005). Alternatively, liquidity shocks can reflect events affecting the overall financial situation of a population of investors: funds experiencing large outflows must sell their holdings, as documented by Coval and Stafford (2007). For regulatory reasons, after large losses, bank must sell risky assets, as discussed by Berndt, Douglas, Duffie, and Ferguson (2005) for the corporate debt market. Khandaniy and Lo (2008) discuss how deteriorating credit portfolios and the need to reduce risk exposure compelled hedge funds to large sales in equity markets in the second week of August 2007, corresponding to a severe liquidity shock.

To analyze the equilibrium reaction of limit order book markets to liquidity shocks we address the following questions: How do prices react and adjust? What is the dynamics of liquidity supply and demand and the corresponding evolution of the order book, trading volume and transactions costs? What are the optimal strategies of the investors and what patterns of order submission do they generate?

We study these issues in a dynamic rational expectations model: Anticipating the dynamics of the order book and trades, agents design their optimal strategies. In equilibrium these strategies give rise to the anticipated market dynamics. We consider an infinite horizon continuous time market with a continuum of rational, risk–neutral competitive investors. Each investor can hold up to one unit of the asset. The asset is in fixed supply and a fraction of the investors is initially endowed with one unit. Investors derive a flow utility from holding the asset. For high-valuation investors this utility flow is greater than for low-valuation investors. To model the aggregate liquidity shock, we follow Duffie, Gârleanu, and Pedersen (2005) and Weill (2007) and assume that at time 0 all investors switch to the low-valuation type. Then, as time goes by, some investors switch back to a high valuation. More precisely, each investor is associated with a Poisson process and switches back to high-valuation at the first jump of this process. Efficiency would require that low-valuation investors would sell to high-valuation investors. Such efficient reallocation of the asset is delayed, however, because all investors are not always present on the market. To model discontinuous market presence, we follow Duffie, Gârleanu, and Pedersen (2005, 2007) and assume that final investors contact the market at Poisson arrival times. The greater the intensity of the Poisson process, the greater the frequency with which investors contact the market.

When contacting the market, investors can place limit orders to sell or buy, and, if they already have orders in the book, they can cancel or modify these. Marketable limit orders (i.e., sell orders at prices lower than or equal to the best bid and buy orders at price greater than or equal to the ask) hit the market quotes and are immediately executed. Non immediately executed limit orders are stored in the book. The dynamics of the order book, in particular the evolution of the bid–ask spread and the depth at the quotes, are endogenous.

In equilibrium, trading occurs in continuous time, but volume, which is initially very low, gradually increases until it reaches a maximum and then progressively dies out. Furthermore, the equilibrium transaction price drops sharply at the time of the liquidity shock and then gradually recovers until it reverts to its long term equilibrium level. The initial price drop and low level of trading are the immediate consequences of the liquidity shock. The hump–shaped pattern of trading volume and progressive recovery of the price reflect the delayed and gradual adjustment of the market due to discontinuous market presence.

High valuation investors, when contacting the market, place buy orders, while low-valuation investors place orders to sell. The reaction of the limit order market to the liquidity shock can be decomposed in two phases. In the first phase, buy orders are placed at very low prices, they set the bid quote and are hit by market orders to sell. But, as time goes by, orders to buy are placed at higher and higher prices. In the second phase, buy orders have reached such high prices that they now hit the ask quotes in the order book. The behavior of the low-valuation investors also varies during the two phases. Initially, they are indifferent between i) placing limit orders to sell at high prices or ii) immediately hitting the bid quote. During this first period their non immediately executed orders are placed at lower and lower ask prices. In the second phase, the low-valuation investors place marketable orders to sell.

Thus, after the shock, there is initially a convergence process, by which the market ask quote declines and the market bid quote increases. Correspondingly, the bid–ask spread declines and the depth on the ask side of the order book grows, starting at high prices and then lower and lower ones. What is the rationale for this pattern? For a low-valuation investor considering how to price her limit sell order, the following tradeoff arises: If she sets a higher price, the benefit is that she gets a better deal. But the cost is that she has to wait longer. The cost of waiting is the time value of money minus the expected utility derived from holding the asset while waiting. Consider a given possible execution time. For early investors, the probability to switch to high valuation before this time is higher than for investors arriving later on the market: thus, at this execution time, the expected utility derived from holding the asset while waiting is higher for early investors than for late investors. Consequently, early investors have a lower cost of waiting and place orders to sell at higher prices than late investors. Hence, the order book progressively fills on the ask side, first at high prices and subsequently at lower and lower prices.

Our theoretical analysis generates several empirical implications in line with stylized facts. Order placement activity concentrates at the best quotes and order of similar types tend to follow each other (in the first phase of our equilibrium there is a sequence of market sell orders while in the second phase there is a sequence of market buy orders.) Both of these implications are in line with the order book and flow dynamics empirically evidenced by Biais, Hillion, and Spatt (1995). The implications of our theoretical model are also in line with the empirical findings of Da and Gao (2007) and Khandaniy and Lo (2008) that after the liquidity shock there is a sharp decline in price and strong order flow imbalance reflecting selling pressure, and then the price gradually recovers.

Progresses in communication technology and the computerization of exchanges and trading rooms have reduced the cost of accessing the market. An important recent development has been algorithmic trading, which increases the speed with which investors can process information and reduces the cost of implementing trading strategies. In the language of our model, this corresponds to an increase in the rate at which investors contact the market. Hendershott, Jones, and Menkveld (2007) offer an empirical study of these developments. Their proxy for algorithmic trading is the ratio of the number of messages (new order placements, cancelations and modifications) to volume traded. They find that this ratio increases, especially when the market becomes more computerized. In the context of our model, this ratio can be analyzed as well as its link to the rate at which traders contact the market. Consistently with the evidence of Hendershott, Jones, and Menkveld, we show that when the rate at which investors can contact the market grows large, so too does the ratio of the number of messages to trading volume. This result stems from two facts. On the one hand, trading volume is bounded above by its Walrasian level, which is finite. On the other hand, as the rate at which investors can contact the market grows large, traders place more demanding orders and cancel and modify them frequently and quickly. This is in line with the stylized fact that on electronic markets, with the progress of algorithmic trading, fleeting orders have become prevalent. Such orders are placed and then very quickly modified or canceled (see Hasbrouck and Saar, 2002).

Our paper is the first to introduce limit orders in the liquidity paradigm initiated by Duffie, Gârleanu, and Pedersen (2005, 2007).[1] By doing so, we add to the rich literature on limit order markets (see the insightful survey by Parlour and Seppi, 2008). The first dynamic models of limit order books have been offered by Foucault (1999) and Parlour (1998). The former proposes an elegant rational expectations model where orders reflect the anticipations of the traders about future market prices, but traders and orders live only one period. The latter presents a rich analysis of the dynamics of depth with long lived orders, but the bid and ask quotes are exogenous. Foucault, Kadan, and Kandel (2005) offer an interesting analysis of long lived orders and endogenous quotes, but only quote improving orders are allowed, while cancelations and modifications are ruled out. We think it is important to propose a model where order placement and cancelations arise in equilibrium, because cancelations have become a very frequent event in electronic markets (see Hasbrouck and Saar, 2002).

Rosu (2009) offers a fully dynamic model, where traders arrive on the market at Poisson

[1] Duffie, Gârleanu, and Pedersen (2002) apply the paradigm to markets for borrowing stocks. Vayanos and Wang (2007), Weill (2008), and Vayanos and Weill (2008) study cross-sectional asset pricing in multi-asset extensions of the original model. Lagos and Rocheteau (2009) and Gârleanu (2008) demonstrate that relaxing the asset holding constraint can have important impact on asset prices and on the entry decisions of intermediaries. Mínguez Afonso (2008) study the externalities involved in investors' entry decisions. Weill (2007) and Lagos and Rocheteau (2009) study the liquidity provision of dealers in response to liquidity shocks similar to the one studied in the present model.

times. The first main difference is that in Rosu, after reaching the market, traders remain continuously present and can modify or cancel orders. Thus, in Rosu's market setup, limit orders are not used to solve the "absent trader problem" identified by Harris (2003). This contrasts with our approach. The second main difference between Rosu's analysis and ours is that he considers strategic traders, while we consider competitive agents: the order placement strategies are driven by current traders' rational expectation of future traders' best replies to their move. No such effects are at play in our model, where, taking as given the equilibrium market dynamics, agents evaluate the costs and benefits of delayed execution and improved price conditions. These various assumptions altogether lead to very different models in Rosu (2009) and in the present paper.

Our model is described in the next section. Section 2 offers a heuristic derivation of the equilibrium. The verification that this is indeed an equilibrium is in Section 3. Section 4 presents implications from our analysis. Section 5 concludes.

1 Model

In this section we present our theoretical limit order market hit by a liquidity shock.

1.1 Asset and agents

Consider the market for an asset, in positive supply $s \in [0, 1)$. The economy operates in continuous time and is populated by a $[0, 1]$ continuum of infinitely lived, competitive and risk-neutral investors who discount the future at the same rate $r > 0$. Investors can hold either zero or one unit of the asset and derive either high or low utility from holding the asset. For high-utility investors the utility flow per unit of time is normalized to $\theta(t) = 1$. For low-utility investors it is equal to $\theta(t) = 1 - \delta$, where $\delta > 0$. There is also a Treasury bill with return r.

At time 0 the market is hit by an aggregate liquidity shock reducing the utility flow to $1 - \delta$ for all investors. But the liquidity shock is transient. Thus, as time goes by, investors randomly switch back to the high utility state, and stay there forever. For simplicity we assume that the times at which investors switch back to high–utility are exponentially distributed, with parameter γ, and independent across investors. Hence the law of large numbers (Sun (2006)) applies and the measure of high-utility investors at time t, denoted by

$\mu_h(t)$, is equal to $1 - e^{-\gamma t}$. That is, the measure of high–utility investors at time t is equal to the probability of being high–utility at that time, conditional on being low–utility at time zero. Because all investors start in the low state we have $\mu_h(0) = 0$.

Conditional on being in the low state at time t, the probability that an individual investor has switched to the high state by time $u \geq t$ is:

$$\pi_h(t, u) = \frac{\mu_h(u) - \mu_h(t)}{1 - \mu_h(t)}. \tag{1}$$

The numerator is the measure of investors who switch from low to high in the interval $[t, u]$, and the denominator is the measure of investors who are still in the low state at time t. Dividing by $(u - t)$ and taking the limit as u goes to t we obtain the hazard rate of switching from low to high at time t, which is equal to γ.

Note that, since

$$s < 1 = \lim_{t \to \infty} \mu_h(t), \tag{2}$$

it follows that, in the steady state, the marginal investor has a high utility. We denote by T_s the time at which the measure of investors with high–utility reaches s:

$$\mu_h(T_s) = s.$$

The evolution of $\mu_h(t)$ and the construction of T_s are illustrated in Figure 1.

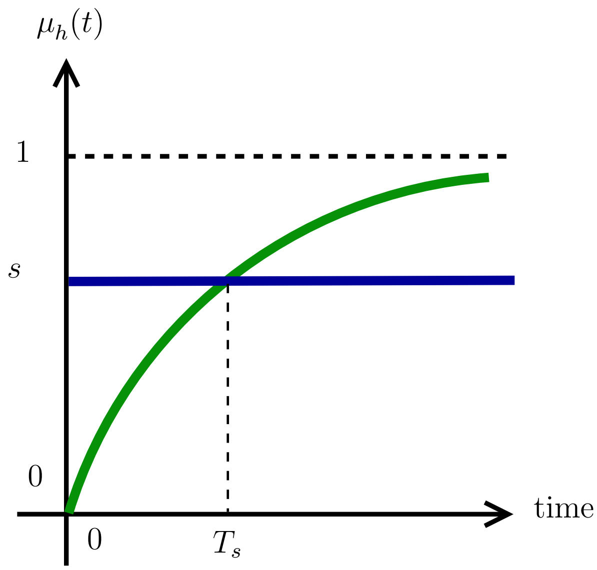

Figure 1: The time path of the fraction $\mu_h(t) = 1 - e^{-\gamma t}$ of high-valuation investors

1.2 Walrasian Equilibrium

First consider the benchmark case where market monitoring is perfect and costless and all investors are permanently ready to trade. The investors are competitive and take the market-clearing price $p(t)$ as given. In equilibrium, $p(t)$ must be such that the marginal investor is indifferent between holding the asset and holding the Treasury bill. After time T_s, the mass of investors who derive high–utility from holding the asset is greater than s. Hence the marginal investor is a high–utility type and the price is:

$$p(t) = \frac{1}{r}.$$

Before time T_s, in contrast, the marginal investor derives low–utility from holding the asset. Hence, the price must be such that,

$$rp(t) = 1 - \delta + \dot{p}(t).$$

This equality ensures that, during a small time interval $[t, t + dt]$, the marginal investor is indifferent between holding the Treasury bill and holding the asset. Indeed, the left–hand–side of the inequality is the instantaneous return on investing $p(t)$ dollars in the Treasury bill. The right–hand–side adds up the marginal-investor utility flow from holding one share, and the capital gain from buying one share at t, and selling it at $t+dt$. The above conditions imply that, at time $t \leq T_s$, the Walrasian price is equal to:

$$p(t) = \frac{1-\delta}{r} + \frac{\delta}{r} e^{-r(T_s - t)}.$$

Thus, the price deterministically increases until it reaches $1/r$ at T_s. This increase reflects the progressive recovery from the aggregate liquidity shock, occuring as investors switch back to high utility flows. The greater the initial liquidity shock (δ) the lower the price. Also, the lower the rate at which agents switch back to high utility (γ), the greater the time it takes for the market to recover (T_s), the lower the price.

In the Walrasian market, trading volume can be readily characterized. Before time T_s, $\mu_h(t) < s$ and all high–utility investors hold one share. Conversely, the only investors who don't hold the asset are low–utility types. Hence there is a mass $1 - s$ of low utility investors who don't own the asset. Trading occurs as these investors switch (at rate γ) to high utility and purchase the asset from low–utility owners. This generates an instantaneous trading volume equal to $\gamma(1-s)dt$. After time T_s, all assets are in the hand of high-valuation investors forever, and the trading volume is zero.

1.3 Trading with imperfect monitoring

Now turn to the case where monitoring is imperfect and costly so that agents are not always trading. Denote by $\rho > 0$ the intensity of market monitoring and assume that investors establish contact with the market at Poisson arrival times with intensity ρ. Contact times are independent across investors and independent from the investors' utility processes. Thus, during each time interval $[t, t+dt]$, a representative cross–section of the population, of size ρdt, contacts the market.

When she contacts the market, an investor trades trough the limit order book. She can place marketable limit buy or sell orders (equivalently referred to as market buy or sell orders) immediately executed at the current ask or bid. Otherwise she can place limit sell (resp. buy) orders at prices above (resp. below) the current market quotes, which are not immediately executed. When contacting the order book, investors can also update and cancel any existing limit order. We assume that order placement, modification or cancelation are costless.

Placing a limit order to sell at some limit price p means selling at that price at the first time such that the market price is greater than or equal to p. This reflects the price priority rule that sell orders at price p should be filled before sell orders at higher prices. The case of limit buy orders is symmetric. Market clearing implies that the number of (market or limit) buy orders executed at time t at the current market price $p(t)$ must be equal to the number of (market or limit) sell orders filled at that price.

2 A heuristic derivation of equilibrium

Our equilibrium derivation is constructive. In section 2.1, we first form a conjecture about general properties of the equilibrium price process and investors' trading strategies. Next, in section 2.2, we use some optimality and market clearing conditions to heuristically refine the initial conjecture: by the end of the section, we obtain a "complete" conjecture for price and strategies. In section 3 we finish the equilibrium derivation by showing that the conjectured trading strategies generate a feasible asset allocation, and that, given the price, the conjectured trading strategies are indeed optimal.

2.1 Price process and strategies: general properties

Our derivation starts with a conjecture of general properties of the price process and investors' trading strategies.

Price process

We conjecture that the price process has the following three properties.

- As in the Walrasian market, since we consider a continuum of agents hit by i.i.d. shocks, the law of large numbers applies, so that aggregate market dynamics is deterministic. We denote the deterministic price process by $p(t)$.

- Since we consider the reaction of the market to a transient adverse liquidity shock, it is natural that the price be increasing with time, i.e., $\dot{p}(t) \geq 0$ (like in the Walrasian case.)

- We also conjecture that the price is continuous. This is a natural conjecture for the following reason: if the increasing price process were to jump up, an investor would strictly prefer to submit a limit order to sell immediately after the jump, rather than a marketable limit order to sell just before the jump. Thus, the asset supply just before the jump would be zero, and the market would not clear. We also conjecture that the price process will be continuously differentiable.

- Ultimately, all agents will have switched to high utility flows and the price will recover its long term value $1/r$. In fact, as in the Walrasian case, price recovery will take place before all agents have switched back: it will occur when the marginal investor in contact with the market has a high–utility. We denote the (endogenous) time at which this occurs by T_f.

Trading strategies

We proceed with a description of investors' trading strategies. The trading strategies of investors map their types into their actions.[2] Thus we need to list the different investors'

[2] While we use the term strategies the reader should bear in mind that the agents in our economy are atomistic and competitive and take the equilibrium price and order book process as given.

types. As in Duffie, Gârleanu, and Pedersen (2005), investors can be categorized according to their utility flow and we denote by "h" those with high utility and "ℓ" those with low utility. In the Walrasian equilibrium, agents with high utility buy from agents with low utility. In the limit order market also, only high utility investors buy, and only low utility investors sell. The difference with the Walrasian case is that they can't do this immediately, and have to wait until they contact the market.

Also as in Duffie, Gârleanu, and Pedersen, investors can be categorized according to whether they own the asset or not (owners are denoted by "o" while non–owners are denoted by "n.") Combining the ownership and utility criteria, let's first consider agents who don't own the asset and have low utility (ℓn). Since they don't own the asset they can't sell it. And, as mentioned above, since they would derive low utility from holding the asset they are not interested in buying it. Hence these agents don't place any orders when they reach the market. Second, turn to the agents who don't own the asset but would derive high–utility from holding it (hn). As mentioned above, these agents seek to buy the asset. Since they rationally anticipate that the price increases with time they have no incentive to delay execution. Hence they place buy orders that are immediately filled.

Before time T_s, the mass of high utility investors, $\mu_h(t)$, is lower than the supply of the asset, s. Hence the instantaneous demand for the asset, stemming from the fraction of high utility investors contacting the market $\rho\mu_h(t)\,dt$, is lower than the supply $\rho s\,dt$, stemming from the fraction of asset owners contacting the market. Thus, all the agents desiring to sell can't trade immediately. Therefore, we conjecture that some of the ℓo investors contacting the market at time $t < T_s$ will find it optimal to place limit orders. In contrast, after time T_s the flow of high utility investors contacting the market exceeds the flow of low utility investors. Hence buy orders are executed either against the current flow of sell orders or against the limit orders to sell stored in the book.

Finally, to complete the analysis of agents who own the asset and contact the market at time t, one needs to take into account the sell orders that these agents may have placed in the book at a previous point in time. Thus, we need to distinguish the following 6 types:

$$\{hn, \ell n, ho, \ell o, hb, \ell b\},$$

where ho and ℓo refer to the agents who own the asset, have not previously placed an order in the book, and have high or low utility respectively, while hb and ℓb denote the investors who own the asset, have previously placed an order in the book, and have high or low utility respectively. Of course, these investors also differ in terms of their limit prices, but we won't need to keep track of these at this stage of the analysis.

In line with the above discussion, we conjecture the trading strategies of investors when contacting the market are the following:

$\underline{hn}$: Investors who don't own the asset and would derive high utility from holding it place immediately executed buy orders and make a transition to type ho.

$\underline{ho}$: Investors who own the asset and derive high utility from their holdings stay put.

$\underline{hb}$: Investors who previously placed a limit sell order in the book but now derive high–utility from holding the asset cancel their previous order and make a transition to type ho.

$\underline{\ell n}$: Investors who would derive low utility from holding the asset but don't own it stay put.

$\underline{\ell o}$: Investors who own the asset and derive low utility from it place sell orders. If their order is immediately executed they make a transition to type ℓn.

$\underline{\ell b}$: Investors who derive low utility from holding the asset, own it and placed a limit order to sell cancel their previous orders and immediately "re-optimize" by placing the same orders as investors of type ℓo.

To complete our description of the trading strategies, we need to spell out the details of the sell orders placed by investors of type ℓo. As discussed above, some of the ℓo investors contacting the market at time $t < T_s$ must place limit orders. Since these agents are all identical, they all place the same limit order. And since the price process is deterministically increasing, choosing the price of the limit order is equivalent to choosing the time at which it will be executed. Denote by $\phi(t)$ the time at which the orders placed at time t are filled. The corresponding price of the limit orders placed at time t is: $p(\phi(t))$. Equilibrium requires that the investors contacting the market at time $t \leq T_s$ be indifferent between selling now at $p(t)$ and selling later at price $p(\phi(t))$. We call $\phi(t)$ the "order placement function" and guess that it is a continuously differentiable, one-to-one mapping from the interval $[0, T_s]$ to the interval $[T_s, T_f]$, where T_f will be determined later. Figure 2 illustrates the position of t and $\phi(t)$. Figure 3 illustrates investors' trading strategies.

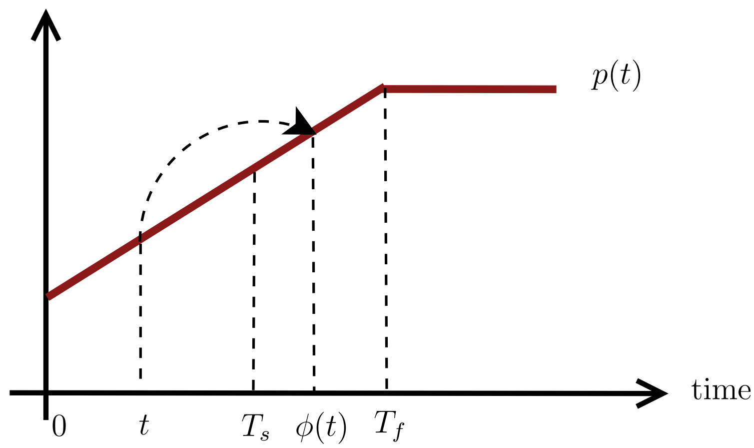

Figure 2: At each time $t < T_s$, ℓo and ℓb investors have two optimal orders. They find it optimal: i) to place a marketable limit order to sell immediately, at time t and also ii) to place a limit order to sell later, at time $\phi(t)$.

2.2 A complete conjecture using optimality and market clearing

So far our conjecture is partial. In this section we use investors' optimality conditions and market clearing conditions to form a complete conjecture about $p(t)$ and $\phi(t)$.

The value of delaying execution and the dynamics of the order book

We start by studying the problem of a low-valuation investor who contacts the market at some time $t \in [0, T_s]$ and seeks to submit a limit order. To determine the best limit order, we fix some execution time $z \in [T_s, T_f]$ and calculate the marginal value of increasing the execution time by dz and behaving optimally thereafter. Note first that the change in execution time is only relevant in the event that the next contact time τ is greater than z. In that event, the increase in execution time has two effects. First, the investor enjoys the asset longer, until $z + dz$, and receives the expected utility

$$\mathbb{E}_t[\theta(z)]dz = (1 - \delta)\, dz + \delta \pi_h(t, z)\, dz, \tag{3}$$

i.e. the investor always enjoys the utility flow $1 - \delta$, but on top of this she may enjoy δ if she has switched to a high utility at some point in the interval $[t, z]$, which happens with probability $\pi_h(t, z)$. The second effect is that the limit order is executed at time $z + dz$

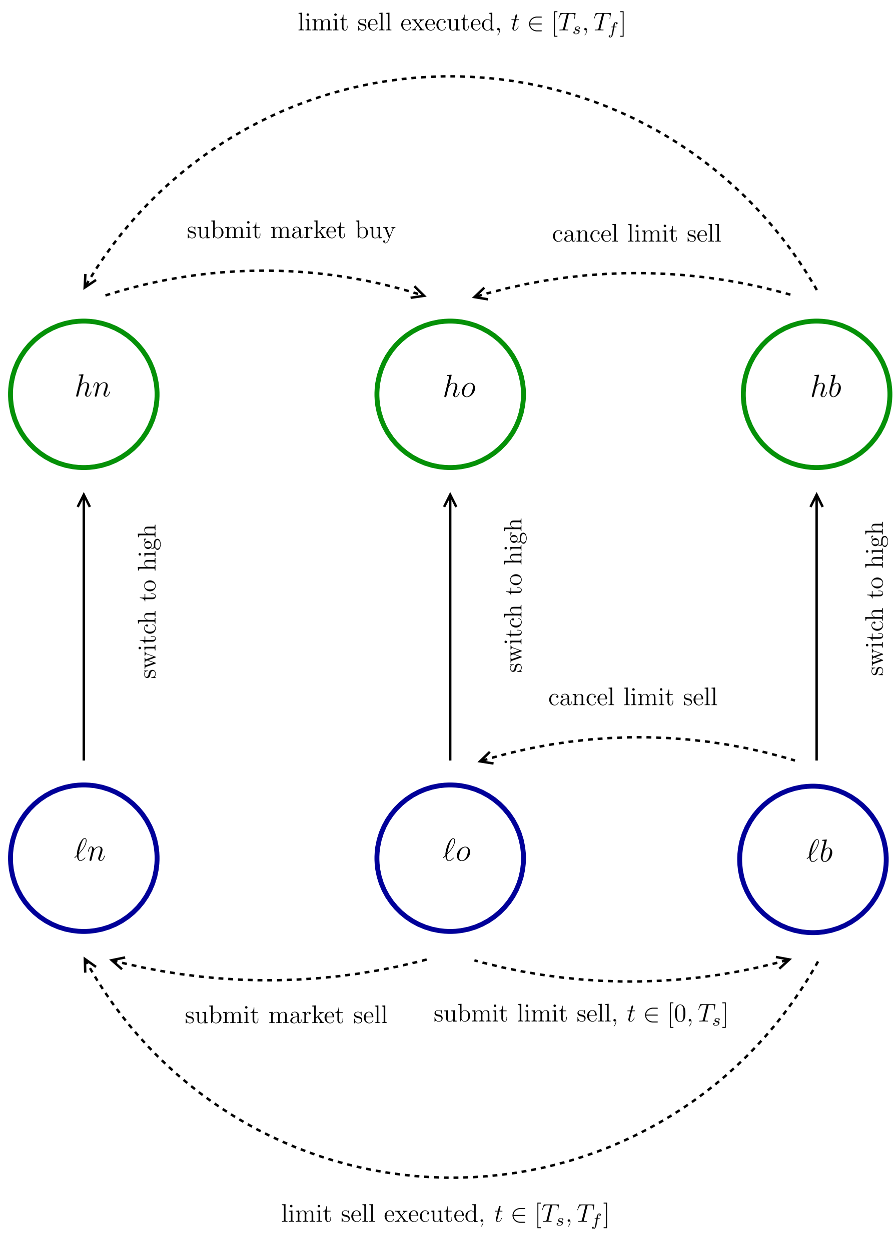

Figure 3: Investors' trading strategies and the associated transitions between types.

instead of time z. The corresponding net utility is

$$\frac{p(z+dz)}{1+rdz} - p(z) = \dot{p}(z)dz - rp(z)dz, \tag{4}$$

after neglecting all second-order terms. The first term on the right-hand side is the capital gain of selling at a later time. The second term is the time cost of delaying the sale. Taking these two effects of equations (3) and (4) together, we obtain that the marginal value of increasing the execution time is equal to:

$$1 - \delta + \delta\pi_h(t, z) + \dot{p}(z) - rp(z). \tag{5}$$

Inspecting (5), one can see that the marginal value of increasing the execution time, from z to $z + dz$, goes up with the probability $\pi_h(t, z)$ that between times t and z there was a switch to high utility. Now, that probability is higher for early t's. Thus, earlier investors are more willing to delay the execution of their sale. This property will imply that, in equilibrium, $\phi(t)$ is decreasing, i.e., investors who contact the market earlier place limit orders that are executed at later times. Equivalently, this means that earlier investors place limit sell orders at higher prices. Thus, the order book fills from the top to the bottom, with successive limit sell orders sequentially undercutting each other.

Implications of optimality for the price process

For the order placement function $\phi(t)$ to be optimal, the marginal value of increasing the execution time beyond $\phi(t)$ must be equal to zero, i.e.:

$$0 = 1 - \delta + \delta\pi_h(t, \phi(t)) + \dot{p}(\phi(t)) - rp(\phi(t)), \tag{6}$$

Using the inverse function $\phi^{-1}(t)$, which is well defined given that $\phi(t)$ is one-to-one, we find that

$$rp(t) = 1 - \delta + \delta\pi_h(\phi^{-1}(t), t) + \dot{p}(t), \tag{7}$$

for all $t \in [T_s, T_f]$. This is the ODE satisfied by the price in $[T_s, T_f]$.

Now turn to the analysis of the price prevailing in the interval $[0, T_s]$. In equilibrium, for all $t \in [0, T_s]$, a low-utility investor must be indifferent between selling his asset outright, with a marketable limit sell order, or submitting a limit order to sell at price $p(\phi(t))$, executed at time $\phi(t)$. To figure out the indifference condition, consider the net utility of submitting a limit order at price $p(\phi(t))$ instead of selling immediately at price $p(t)$.

- First, consider what happens at time t. If the investor submits a limit order she receives nothing. If she submit a marketable limit order she receives $p(t)$. So, the net utility at time t of submitting the limit order instead of selling immediately is equal to

$$- p(t). \tag{8}$$

- Second, consider what happens until the execution time $\phi(t)$, or the next contact time, τ, whichever comes first. If the investor submits a limit order, she enjoys a flow utility $\theta(u)$. If she submits a marketable limit order, she does not own the asset so the utility flow is zero. After discounting and taking expectations, we find that, over the interval $(t, \min\{\tau, \phi(t)\})$, the net utility of submitting a limit order rather than selling is:

$$\mathbb{E}_t\left[\int_t^{\min\{\tau,\phi(t)\}} e^{-r(u-t)}\theta(u)\,du\right]. \tag{9}$$

- Third, consider what happens at the next contact time, τ, or at the execution time $\phi(t)$, whichever comes first. There are two possible scenarii:

  - The first scenario corresponds to the case where $\tau < \phi(t)$. Then, if the investor has a limit order, an optimal strategy is to cancel the order, sell the asset at price $p(\tau)$, and behave as a non-owner, i.e. as if she had in fact sold her asset at time t. Thus the net utility at time τ of having placed a limit order at time t is simply $p(\tau)$, the value of canceling the order and selling the asset.
  - The second scenario corresponds to the case where $\phi(t) \leq \tau$: the asset is sold if the investor has a limit order outstanding, and nothing happens otherwise. Thus, the net utility at time $\phi(t)$ of having placed a limit order at time t is equal to $p(\phi(t))$.

  Taken together, we obtain the following expression for the net utility at the next contact time or at the execution time, whichever come first:

$$\mathbb{E}_t\left[e^{-r(\min\{\tau,\phi(t)\}-t)}p(\min\{\tau,\phi(t)\})\right] \tag{10}$$

Collecting equations (8), (9) and (10), we find that the total net utility of placing the limit order rather than selling immediately is equal to

$$-p(t) + \mathbb{E}_t\left[\int_t^{\min\{\tau,\phi(t)\}} e^{-r(u-t)}\theta(u)\,du + e^{-r(\min\{\tau,\phi(t)\}-t)}p(\min\{\tau,\phi(t)\})\right].$$

This must be equal to zero for the agent to be indifferent. So the price must be equal to

$$p(t) = \mathbb{E}_t \left[\int_t^{\min\{\tau,\phi(t)\}} e^{-r(u-t)}\theta(u)\, du + e^{-r(\min\{\tau,\phi(t)\}-t)} p(\min\{\tau, \phi(t)\}) \right]. \qquad (11)$$

The present value formula above is fairly intuitive: the price adjusts so that the low-valuation investor is indifferent between selling now or holding the asset and selling it at the next contact time, or at the execution time, whichever comes first.

Together with the condition that $p(t) = 1/r$ for all $t \geq T_f$, the ODE (7) and the equation (11) will completely determine the price path.

2.3 Market clearing and the order placement function

In the previous section we have determined our conjecture for $p(t)$, given any decreasing order-placement function $\phi(t)$. We now proceed to show heuristically that, given our conjectured trading strategy, there is a unique decreasing $\phi(t)$ that is consistent with market clearing.

Consider any interval of time $[t, t + dt]$ before T_s. On one hand, the number of assets brought to the market by investors of all types is equal to $\rho s\, dt$. On the other hand, the number of high-valuation investors who contact the market is equal to $\rho\mu_h(t)\, dt$. According to our conjectured trading strategies, all of these high-valuation investors walk out of the market with one unit of the asset.[3] Thus, after allocating one asset to each high utility investors contacting the market, there remains a net quantity

$$\rho\, (s - \mu_h(t))\ dt > 0$$

of assets that, in an equilibrium, have to be held by low utility investors. Since these low-utility investors hold on their asset, they can't be placing market orders. Instead, they place limit orders.

These limit orders, submitted during the interval $[t, t+dt]$, will be either executed during the time interval

$$[\phi(t + dt), \phi(t)] = [\phi(t) + \phi'(t)\, dt, \phi(t)],$$

or canceled before $\phi(t)$. To compute the probability that the order is still outstanding by the

[3] That is, if they are of type ho, they hold on to their asset; and if they are of type hn, they buy one unit of the asset and become ho.

execution time $\phi(t)$, recall that an investor will always find it optimal to cancel (and maybe resubmit a modified order) at his next contact time with the market. Thus, the probability that the order is still outstanding is simply equal to the probability

$$e^{-\rho(\phi(t)-t)} \tag{12}$$

that the investor *does not contact* the market during $[t, \phi(t)]$. By the law of large numbers, this probability also represents the fraction of limit orders submitted at time t which are still outstanding at time $\phi(t)$. Thus, according to our conjectured trading strategies, the number of limit orders executed during $[\phi(t+dt), \phi(t)] = [\phi(t) + \phi'(t)\, dt, \phi(t)]$, must be equal to

$$\rho\,(s - \mu_h(t))\; dt \times e^{-\rho(\phi(t)-t)}. \tag{13}$$

Equivalently this is the number of limit orders executed at price $p(\phi(t))$.

Now consider the market buy orders that will be executed against these limit sell orders during that time interval. Investors bring a flow $\rho s\,|\phi'(t)|\, dt$ of assets to the market. The number of high-utility investors who contact the market is equal to $\rho\mu_h(\phi(t))|\phi'(t)|\, dt$. According to our conjectured trading strategies, all of these investors walk out of the market with one unit of the asset. Some of this demand is matched by the supply $\rho s|\phi'(t)|\, dt$ that investors bring to the market during the interval $[\phi(t) + \phi'(t)\, dt, \phi(t)]$. The remainder,

$$\rho\,(\mu_h(\phi(t)) - s)\; |\phi'(t)|\, dt. \tag{14}$$

has to be matched by the supply (13) of limit orders. Equating (14) with (13), we obtain the market clearing condition:

$$\rho\,(\mu_h(\phi(t)) - s)\,|\phi'(t)| = \rho\,(s - \mu_h(t))\, e^{-\rho(\phi(t)-t)}.$$

Multiplying both sides by $e^{\rho\phi(t)}$, integrating over $[t, T_s]$, and keeping in mind that $\phi'(t) < 0$ because $\phi(t)$ is decreasing, we obtain

$$\begin{aligned} &\int_t^{T_s} \rho\,(\mu_h(\phi(z)) - s)\; |\phi'(z)| e^{\rho\phi(z)}\, dz = \int_t^{T_s} \rho\,(s - \mu_h(z))\, e^{\rho z}\, dz \\ \Leftrightarrow\quad &\int_t^{\phi(t)} \rho\,(s - \mu_h(z))\, e^{\rho z}\, dz = 0, \end{aligned} \tag{15}$$

where the second line follows from the change of variable $u = \phi(z)$. Note that this equation also pins down the first time T_f at which the limit order book is empty:

$$\int_0^{T_f} \rho\left(s - \mu_h(z)\right) e^{\rho z}\, dz = 0. \tag{16}$$

Equipped with these two equations, we can show:

Proposition 1 (Order Placement)**.** *There exists a unique function $\phi(t)$ such that (15) and (16) hold. The function $\phi(t)$ is continuously differentiable and decreasing, it maps $[0, T_s]$ onto $[T_s, T_f]$. In addition, $\phi(t)$ is increasing in s and decreasing in ρ and γ.*

The intuition for the above comparative statics with respect to s, ρ and γ is the following. Limit orders to sell placed at time $t < T_s$ are executed at time $\phi(t) > T_s$, where T_s is the time at which sufficiently many investors have switched to high utility for the mass of high utility agents to exceed the supply. As s goes up, this time goes up, and so does $\phi(t)$, which is greater than T_s. In other words, as the supply of the asset goes up, it takes longer for the market as a whole to absorb the liquidity shock, and correspondingly it takes longer for limit sell orders to be executed. In contrast, when the rate at which agents switch to high utility (γ) or the rate at which they contact the market (ρ) increase, the market absorbs the liquidity shock faster, and correspondingly limit sell orders get executed faster.

3 Equilibrium verification

In this section, we complete our construction by showing: i) that the asset allocation generated by the trading strategy is feasible, ii) that the price function is increasing and continuously differentiable, and iii) that, given the price process, the conjectured trading strategies are optimal.

3.1 Feasibility

Appendix A.2 explicitly solves for the population dynamics, i.e., the measure of each investor's type at each point in time, when investors follow the conjectured trading strategies. The calculations confirm that supply expressed at the current market price is equal to current demand. Thus we can state the following proposition:

Proposition 2. *The asset allocation implied by the conjectured trading strategy and the order placement function is feasible: supply expressed at the current market price is equal to current demand.*

3.2 Monotonicity and differentiability

We now verify our conjectures about the price path. After substituting the order placement function $\phi(t)$ into the price equations (11) and (7), the following proposition obtains:

Proposition 3 (Differentiability and monotonicity). *The price process is strictly increasing and continuously differentiable over $[0, T_f)$. Before T_f it solves the ODEs:*

$$\begin{aligned} t \in [0, T_s], \quad rp(t) &= 1 - \delta - \delta \int_t^{\phi(t)} \frac{\partial \pi_h}{\partial t}(t, u) e^{-(r+\rho)(u-t)} \, du + \dot{p}(t) \\ t \in [T_s, T_f], \quad rp(t) &= 1 - \delta + \delta \pi_h(\phi^{-1}(t), t) + \dot{p}(t), \end{aligned}$$

and, for $t \geq T_s$, $p(t) = 1/r$.

The proposition highlights that, before T_f, the dynamics of the price reflects the probability that the marginal holder of the asset can switch to high utility.

3.3 Optimality

The only thing that remains to be verified is that, given the price, the conjectured trading strategies are optimal.

No limit buy orders

The first step is a simple Lemma allowing us to rule out from start limit buy orders:

Lemma 1 (No limit buys). *Limit order to buy are sub-optimal for all investors.*

The reason is that the price process is increasing. This implies that a limit order to buy at price $p < p(t)$ is never executed, so submitting such an order is weakly dominated by staying put. On the other hand, a limit order to buy at price $p > p(t)$ is immediately executed at price p: clearly, an investor is strictly better off submitting a market order to buy, which is executed at the lower price $p(t)$.

Optimality verification lemmas

Next, we apply the the Principle of Optimality of dynamic programming and check "one-stage" deviations. That is, at every time and for all types, we check that an investor is worse off if i) she deviates from the conjectured strategy once when contacting the market at time $t \geq 0$ and ii) follows the candidate optimal strategy thereafter.

We let $V_{\ell o}(t)$, $V_{\ell n}(t)$, $V_{ho}(t)$ and $V_{hn}(t)$ be the respective continuation utilities at time t of a ℓo, ℓn, ho, hn investors who behave according to our conjecture trading strategies. Investors with outstanding limit orders have to be distinguished according to the execution time of their order: we let $V_{\ell b}(t, z)$ and $V_{hb}(t, z)$ be the continuation utility of low- and high- valuation investors, respectively, with a limit order to be executed at time $z \geq t$, who behave according to our conjectured trading strategies. Note that since the price is strictly increasing over $[0, T_f]$ and constant afterwards, the only feasible execution times at time t are $z \in [t, T_f]$. In our proof we study a "relaxed" problem whereby investors are allowed to submit *any* execution time $z \in [t, \infty)$. Clearly, if the conjectured trading strategies solves the relaxed problem, they also solve the original problem. One advantage of the relaxed problem is that all possible sell orders can be represented by their execution time: placing a market sell order correspond to the execution time $z = 0$, placing a limit sell orders to execution times $z \in [t, T_f]$, and placing no order is payoff equivalent to choosing an execution time $z \to \infty$.

Lemma 2 (Bellman Principle for Low Valuations)**.** *Given some contact time $t \geq 0$, consider the problem of choosing an execution time $z \geq t$ in order to maximize:*

$$V_{\ell b}(t, z) - V_{\ell n}(t) - p(t). \tag{17}$$

The conjectured trading strategies of low-valuation investors are optimal if the maximum is zero and is achieved at $z = t$ and $z = \phi(t)$ for all $t < T_s$, and at $z = t$ for all $t \geq T_s$.

As an illustration, consider the case of an ℓo investor contacting the market at some time $t < T_s$. The conjectured trading strategy is to submit a market order to sell, with a continuation utility $V_{\ell n}(t) + p(t)$. The possible deviations are to either submit limit order to sell at time $z \in [t, \infty)$, with a continuation utility $V_{\ell b}(t, z)$ or to stay put, with a continuation utility $V_{\ell o}(t) = \lim_{z \to \infty} V_{\ell b}(t, z)$. Clearly, the conjectured trading strategy is optimal if:

$$V_{\ell b}(t, z) \leq V_{\ell n} + p(t),$$

with an equality when $z = t$ and $z = \phi(t)$. That is, for a ℓo investor, the conditions of the above lemma are indeed sufficient for optimality. Other low-valuation investors can be studied similarly.

Now turn let us to the case of high-valuation investor:

Lemma 3 (Bellman Principle for High Valuations)**.** *Given some contact time $t \geq 0$, consider the problem of choosing an execution time $z \geq t$ in order to maximize:*

$$V_{hb}(t, z) - V_{ho}. \tag{18}$$

Then, the conjectured trading strategies of low-valuation investor are optimal if i) the supremum is zero; and ii) the supremum is achieved as $z \to \infty$.

Note that $V_{ho} = 1/r$ is a constant function of time because *ho* investors hold on the asset forever. The Lemma is proved in Appendix A.5. To illustrate it, consider the case of a *ho* investor who contacts the market at time t. The prescribed trading strategy is to stay put, with a continuation utility is $V_{ho}(t) = \lim_{z \to \infty} V_{hb}(t, z)$. The possible deviations are to submit limit orders to sell at $z \geq t$. Thus, the conjectured trading strategy is optimal if

$$V_{hb}(t, z) - V_{ho} \leq 0,$$

with an equality in the limit $z \to \infty$.

Optimality verification

Equipped with this to Lemmas, one obtains the following result:

Proposition 4. *Given the conjectured price process, the conjectured trading strategies are optimal and, during the interval $[0, T_f]$, they achieve a strict maximum.*

The second part of the Proposition means in particular that the timing of limit-order submission is determinate. Namely, a ℓo investor who contacts the market at time t has exactly two optimal orders, selling at time t or submitting a limit order to sell at price $p(\phi(t))$. All other orders would result in a utility loss.

Taken together Propositions 2 and 4 imply the following result:

Theorem 1. *The conjectured price and trading strategies constitute a limit-order equilibrium.*

4 Implications

4.1 Trading volume

During the time interval $[0, T_s]$, the dynamics of the measure of hn investors is the following

$$\dot{\mu}_{hn}(t) = -\rho\mu_{hn}(t) + \gamma\mu_{\ell n}(t). \tag{19}$$

where: $\mu_{hn}(t)$ and $\mu_{\ell n}(t)$ denote the measure of hn and ℓn investors, $\dot{\mu}_{hn}(t) = d/dt[\mu_{hn}(t)]$. The first term on the right-hand side of (19) arises because, during a small time interval, there is a flow $\rho\mu_{hn}dt$ of hn investors who contact the market. All of them submit marketable limit orders to buy and become ho investor. The second term arises because there is a flow $\gamma\mu_{\ell n}dt$ of ℓn investors who switch to a high utility.

Since $\mu_{hn}(t) + \mu_{\ell n}(t) = 1 - s$, we have that $\mu_{\ell n} = 1 - s - \mu_{hn}(t)$ which, together with (19) implies that

$$\dot{\mu}_{hn}(t) = -(\rho + \gamma)\mu_{hn}(t) + \gamma(1 - s).$$

With the initial condition that $\mu_{hn}(0) = 0$, this gives

$$\mu_{hn}(t) = \frac{\gamma(1-s)}{\rho+\gamma}\left(1 - e^{-(\rho+\gamma)t}\right).$$

During the interval $[t, t + dt]$ a fraction ρdt of the hn investors contact the market. Hence, before T_s, instantaneous trading volume is equal to:

$$V(t) = \rho\mu_{hn}(t) = \gamma(1-s)\frac{\rho}{\rho+\gamma}\left(1 - e^{-(\rho+\gamma)t}\right) \tag{20}$$

Equation (20) shows that trading volume with imperfect monitoring is lower than its Walrasian counterpart ($\gamma(1 - s)$).

To study how the intensity of market monitoring affects trading volume, differentiate V in (20) with respect to ρ. This yields

$$\frac{\partial V}{\partial \rho}(t) = \frac{\gamma(1-s)}{(\rho+\gamma)^2}e^{-(\rho+\gamma)t}(\gamma e^{(\rho+\gamma)t} - \gamma + \rho(\rho+\gamma)t),$$

which is positive. Thus, trading volume increases with the intensity of market monitoring. This is intuitive. If high-utility investors can contact the market more often, trading volume goes up. Note also that, as ρ goes to infinity, volume goes to $\gamma(1-s)$, which is the trading volume in the Walrasian market.

Since (20) is symmetric in ρ and γ we also have that trading volume increases in the rate at which low utility investors switch to high utility. Indeed, in this model it is the buyer side of the market which constrains trading, so if there are more high utility investors eager to buy, there is more trade. An increase in γ generates an increase in flow of new high utility investors.

On the other hand, inspecting (20), one can see that volume goes down with s. This is similar to what happened in the Walrasian case and arises because trading volume reflects the imbalance between the number of high utility agents and the number of assets, s: an increase in s leads to a decrease in this imbalance, and a decrease in volume.

4.2 The number of orders in the book

Let $L(t,v)$ denote the stock of limit orders submitted after time t, still in the book at time $v \in [t, \phi(t)]$. We have

$$\frac{\partial L}{\partial v}(t,v) = -\rho L(t,v) + \rho\left(s - \mu_h(v)\right).$$

The first term is the flow of cancelations. For $v \in [t, T_s]$ the second term is the positive flow of new limit orders placed in the book. For $v \in [T_s, \phi(t)]$, it is the negative flow of limit orders leaving the book because they are executed. Recall that the orders leaving the book during $[T_s, \phi(t)]$ were indeed submitted during $[t, T_s]$. Integrating this ODE with the initial condition $L(t,t) = 0$, we find that:[4]

$$L(t,v) = \int_t^v \rho\left(s - \mu_h(z)\right) e^{-\rho(v-z)}\, dz.$$

Setting $t = 0$, this yields the stock of limit orders in the book at time v:

$$L(0,v) = \int_0^v \rho\left(s - \mu_h(z)\right) e^{-\rho(v-z)}\, dz. \tag{21}$$

This equation has an intuitive interpretation. The stock of limit orders in the book at time v is the cumulated net flow of orders placed in the book, $\rho\left(s - \mu_h(z)\right)$, multiplied by

[4]Note that the implicit equation for $\phi(t)$ is simply $L(t, \phi(t)) = 0$.

the proportion that these orders have not been canceled at time v (which is similar to the proportion of orders not canceled given in equation (12)).

Substituting the value of $\mu_h(t)$ in equation (21), we get:

$$L(0, v) = \int_0^v \rho \left(s - 1 + e^{-\gamma z}\right) e^{-\rho(v-z)} \, dz. \tag{22}$$

So the derivative of $L(0, v)$ with respect to γ is:

$$\int_0^v -z e^{-\gamma z} e^{-\rho(v-z)} \, dz < 0. \tag{23}$$

Thus, the smaller the rate γ at which investors switch back to high utility, the greater the need to use limit orders to wait for counterparties, the larger the number of orders accumulated in the book.

4.3 How the limit order market absorbs the liquidity shock

As discussed in Section 3, before time T_s, the flow of low utility investors is greater than the flow of high utility investors. We interpret this as a buyers' market. In this context, low utility investors who own the asset are indifferent between placing limit order to sell and market orders to sell. These market orders are immediately executed, at the current market price, against the flow of orders to buy. The latter can be interpreted as marketable orders to buy, setting the bid price, which is also the current transaction price.

After time T_s, in contrast, the flow of low utility investors is lower than the flow of high utility investors. We interpret this as a sellers' market. High utility investors buy at the limit selling price established by previously placed orders, i.e., the buyers hit the ask quote.

Thus there are two market regimes: before T_s, there is a buyer market, in which market orders to sell repeatedly hit the bid quote, while after T_s there is a seller market, in which market orders to buy repeatedly hit the ask quote. And, during the first phase, there is a sequence of new limit orders to sell placed within the best quotes and undercutting each other. These patterns are consistent with the stylized facts observed in limit order markets, in particular the fact that similar order types tend to follow each other (see Biais, Hillion, and Spatt, 1995; Griffiths, Smith, Turnbull, and White, 2000; Ellul, Holden, Jain, and Jennings, 2007).

Our previous results also have implications for the dynamics of the spread and the book during these two regimes. They are illustrated in Figure 4. As can be seen in the figure, just after the liquidity shock the spread is large. Then, limit orders to sell accumulate in the book, driving down the ask quote. And marketable limit buy orders are placed at higher and higher level. This results in a decrease in the bid–ask spread. Also, as can be seen in the figure, the number of orders in the book is very low just after the shock. But, as new limit orders to sell are placed in the book, depth progressively builds up. Yet, at some point, cancelations and execution of market buy orders lead to a decrease in the stock of limit orders in the book.

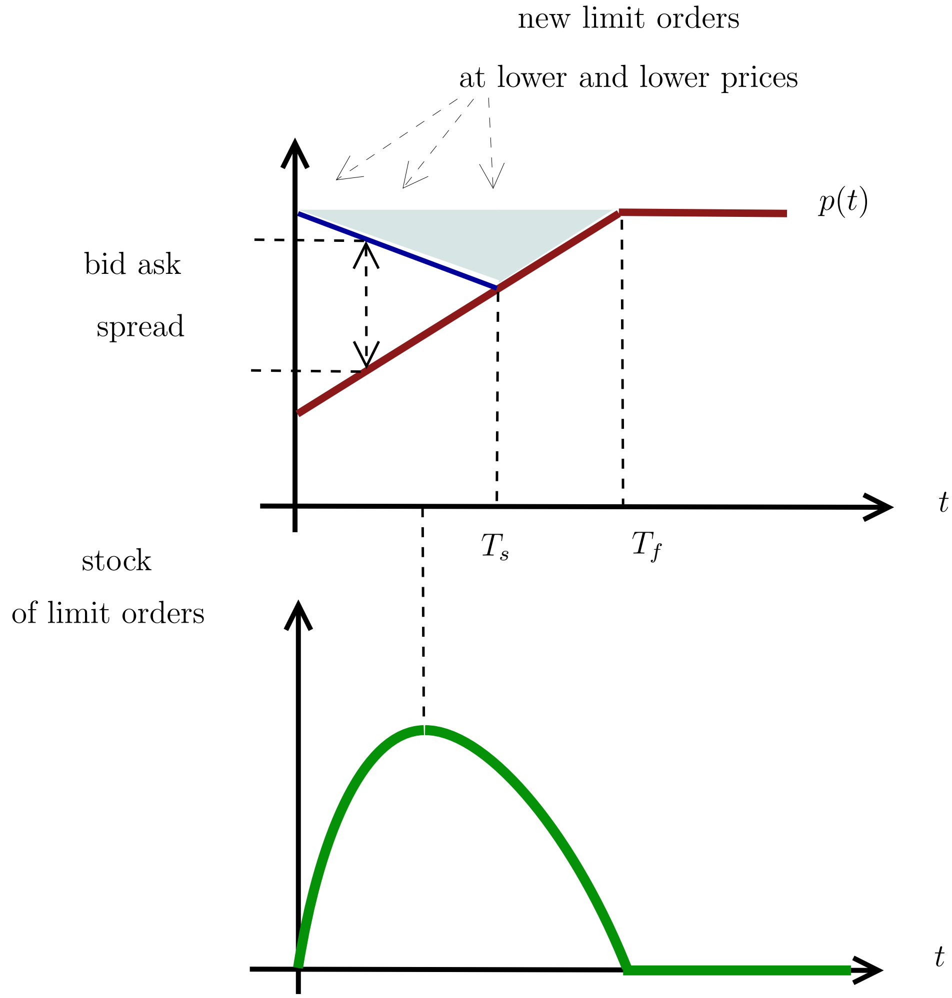

Figure 4: Price and order book dynamics

4.4 Technological Change

During the last 20 years, exchange trading technology has improved dramatically. The ability for investors to observe market quotes and trades and rapidly place orders has expanded.

Agents increasingly rely on computers to collect and proces information, generate alerts on market movements and inform trading and investment decisions. An extreme and important form of the development of such computerization has been the growth of algorithmic trading.

Hendershott, Jones, and Menkveld (2007) offer interesting evidence on these issues. They proxy algorithmic trading by the ratio of the number of new orders, modifications and cancelations (i.e., messages) to trading volume. The idea is that, without algorithmic trading, investors will use a few large orders, while with algorithmic trading they will split up these orders in several smaller ones and often cancel and revise these orders. Hendershott, Jones, and Menkveld also take advantage of the fact that, during the period they study, the NYSE progressively implemented its "autoquote" system, which facilitates the placement of electronic orders, and thus algorithmic trading. They find that, as "autoquote" gets implemented, the proxy for algorithmic trading (i.e., the ratio of messages to volume) goes up.

Our analysis offers a framework to shed light on these evolutions. The growth of algorithmic trading and exchange computerization correspond to an increase in the speed with which agents contact the market, i.e., in our model an increase in ρ. For simplicity, in this subsection, we focus on the case where ρ goes to infinity, i.e., the market approaches the continuous trading Walrasian benchmark. Our first result is:

Proposition 5. *For each $t \in (0, T_s]$, as ρ goes to infinity, the number of orders in the book at time t converges to*

$$\lim_{\rho \to \infty} L(t) = L^\infty(t) = s - \mu_h(t).$$

The proof is in the appendix. Since, in the limit $\mu_{ho}(t) = \mu_h(t)$, it follows that $L^\infty(t)$ is equal to the number of assets in the hand of low-utility investors. Therefore, in the limit, although agents can effectively trade continuously, the limit order book is *not* empty. Intuitively, low-valuation investors who choose not to trade now always post a limit order, because there remains a remote chance that they are not able to re-contact the market very quickly. Correspondingly, since orders in the book are associated to limit prices greater than $p(T_s) > p(t)$, the bid-ask spread at time $t < T_s$ converges to some non-zero limit. Now turn to the behavior of cancelations:

Proposition 6. *For each $t \in (0, T_s)$, as ρ goes to infinity, the flow of cancelations goes to infinity. Moreover, it is strictly increasing in ρ, for ρ large enough.*

The intuition for this result is the following: at any time $t < T_s$, the flow of cancelation is equal to

$$C(t) = \rho L(t).$$

Thus, since the book does not become empty as ρ goes to infinity, the flow of cancelations goes to infinity. The proof of monotonicity is in the appendix. Finally consider the flow of messages and its relation to trading volume, the statistics that is studied empirically by Hendershott, Jones, and Menkveld (2007):

Proposition 7. *For each $t \in (0, T_s)$, as ρ goes to infinity, the ratio of messages to volume goes to infinity. Moreover, it is strictly increasing in ρ, for ρ large enough.*

The flow of messages at time t is

$$M(t) = \rho\mu_{\ell o}(t) + 2\rho\mu_{\ell b}(t) + \rho\mu_{hb}(t) + \rho\mu_{hn}(t),$$

which is the sum of four components:

- The flow $\rho\mu_{\ell o}(t)$ of ℓo investors contacting the market, whose message is an order to sell.
- Twice the flow of $\rho\mu_{\ell b}(t)$ of ℓb investors who contact the market, because these agents send *two* messages: they cancel their order and submit another one.
- The flow $\rho\mu_{hb}(t)$ of hb investors contacting the market, whose message is to cancel their limit sell order.
- The flow $\rho\mu_{hn}(t)$ of hn investors contacting the market, whose message is to submit a market order to buy.

Rearranging, we obtain:

$$\begin{aligned}
M(t) &= \rho\Big(\mu_{\ell o}(t)+\mu_{\ell b}(t)-\mu_{hn}(t)\Big)+\rho\Big(\mu_{\ell b}(t)+\mu_{hb}(t)\Big)+2\rho\mu_{hn}(t)\\
&= \rho\Big(\mu_{\ell o}(t)+\mu_{\ell b}(t)+\mu_{\ell n}(t)-\mu_{\ell n}(t)-\mu_{hn}(t)\Big)+\rho L(t)+2V(t)\\
&= \rho\Big(1-\mu_h(t)-(1-s)\Big)+\rho L(t)+2V(t)\\
&= \rho\Big(s-\mu_h(t)\Big)+\rho L(t)+2V(t).\\
&= \rho\Big(s-\mu_h(t)\Big)+C(t)+2V(t).
\end{aligned}$$

In words, the flow of messages is the sum of the flow $\rho(s-\mu_h(t))$ of new limit orders, the flow $C(t)$ of cancelations, and twice the volume. One sees that, although the volume is bounded, the number of messages goes to infinity as ρ goes to infinity. The ratio of messages to volume is equal to

$$\frac{M(t)}{V(t)}=2+\frac{\rho\Big(s-\mu_h(t)\Big)+C(t)}{V(t)}$$

The proof that this ratio increases in ρ is in the appendix.

5 Conclusion

This paper offers a continuous time model of order book dynamics, in which the arrival of traders is random. The order flow, including the placement of limit orders, cancelations & modifications, as well as the dynamics of prices and trading volume, are endogenous. We find that, after a liquidity shock there are two phases. First, there is a "buyers' market," in which the flow of sell orders exceeds the flow of buy orders and trades hit the bid quote. During that phase, the bid-ask spread is initially high, but progressively tightens, while, in parallel the depth in the book builds up. Second, there is a "seller's market" during which the flow of buy orders exceeds the flow of buy orders, and trades hit the ask side of the book. The dynamics generated by our model match the stylized facts on order books, with clustering of the activity at the best quotes, undercutting and serial correlation in order types. Our model also sheds light on the consequences of the increase computerization of markets & the growth of algorithmic trading. Our analysis implies that these changes imply an increase in trading volume and an even stronger increase in message traffic, corresponding to frequent cancelations and modifications. These results also corroborate empirical evidence.

In further research we plan to use this framework to study the welfare properties of limit order markets, the endogenization of trading technology, and the design of markets.

A Proofs

A.1 Proof of Proposition 1

Consider some time $t \leq T_s$. Then, the order execution time, $\phi(t)$, is a solution $v \in [T_s, \infty)$ of the equation $H(t,v) = 0$, where

$$H(t,v) \quad = \quad \int_t^v (s - \mu_h(u))\, e^{\rho u}\, du \equiv \int_t^v h(u)\, du \tag{24}$$

and $h(u) \equiv \rho\, (s - \mu_h(u))\, e^{\rho u}$.

Existence and uniqueness. Because $s - \mu_h(u)$ is strictly decreasing in u and is equal to zero when $u = T_s$, it follows that $H(t, T_s) \geq 0$, that $H(t,v)$ is decreasing in $v \geq T_s$, and goes to minus infinity as $v \to \infty$. Taken together, this implies that, for all $t \in [0, T_s]$, there is a unique $\phi(t) \geq T_s$ such that $H(t, \phi(t)) = 0$. We then let define $T_f \equiv \phi(0)$.

Monotonicity. For all $t < T_s$, $H(t, T_s) > 0$ and thus $\phi(t) > T_s$. Therefore $\partial H/\partial v(t, \phi(t)) = h(\phi(t)) < 0$. Next, an application of the Implicit Function Theorem shows that the function $\phi(t)$ is continuously differentiable over $(0, T_s)$, with a derivative that is equal to:

$$\phi'(t) = \frac{h(t)}{h(\phi(t))}. \tag{25}$$

Note that because $t < T_s < \phi(t)$, we have that $h(t) > 0$ and $h(\phi(t)) < 0$, so $\phi'(t) < 0$. Clearly, $\phi'(t)$ can be extended by continuity at $t = 0$. The last thing to establish is that $\phi(t)$ is continuously differentiable at $t = T_s$. We start by showing that it is differentiable. First, we apply Taylor Theorem, up to the second order, and we obtain:

$$H(t, \phi(t)) \quad = \quad \int_t^{T_s} h(u)\, du + \int_{T_s}^{\phi(t)} h(u)\, du = -\frac{(t - T_s)^2}{2} h'(\xi_t) + \frac{(\phi(t) - T_s)^2}{2} h'(\psi_t),$$

where $\xi_t \in [t, T_s]$ and $\psi_t \in [T_s, \phi(t)]$. Since $H(t, \phi(t)) = 0$, and $t < T_s < \phi(t)$, solving this equation gives

$$\frac{\phi(t) - T_s}{t - T_s} = -\sqrt{\frac{h'(\xi_t)}{h'(\psi_t)}},$$

which goes to -1 as t goes to T_s because both ξ_t and ψ_t go to T_s and $h'(T_s) \neq 0$. It thus follows that $\phi(t)$ is differentiable at $t = T_s$, with $\phi'(T_s) = -1$. For continuous differentiability, we can write:

$$\phi'(t) \quad = \quad \frac{h(t)}{h(\phi(t))} = \frac{h(t) - h(T_s)}{t - T_s} \times \frac{\phi(t) - \phi(T_s)}{h(\phi(t)) - h(\phi(T_s))} \times \frac{\phi(t) - \phi(T_s)}{t - T_s},$$

because $h(T_s) = 0$ and $\phi(T_s) = T_s$. Letting t go to T_s, we find that

$$\lim_{t \to T_s} \phi'(t) = \frac{h'(T_s)}{h'(\phi(T_s))} \times \lim_{t \to T_s} \frac{\phi(t) - \phi(T_s)}{t - T_s} = \phi'(T_s),$$

keeping in mind that $\phi(T_s) = T_s$.

The order placement function is increasing in s. We consider $t < T_s$. Because $\phi(t) > T_s$, it follows that $\partial H/\partial v = \rho\,(s - \mu_h(v))\,e^{\rho v} < 0$. Also, taking partial derivatives with respect to s, we obtain $\partial H/\partial s = \int_t^v e^{\rho u}\,du > 0$. Taken together with an application of the Implicit Function Theorem, these imply that $\phi(t)$ is increasing in s.

The order placement function is decreasing in ρ. Taking partial derivative with respect to ρ, evaluated $\phi(t)$

$$\begin{aligned}
\frac{\partial H}{\partial \rho}(t, \phi(t)) &= \int_t^{\phi(t)} (s - \mu_h(u))\,e^{\rho u}\,du + \int_t^{\phi(t)} \rho u\,(s - \mu_h(u))\,e^{\rho u}\,du \\
&= H(t, \phi(t))/\rho + \int_t^{\phi(t)} \rho u\,(s - \mu_h(u))\,e^{\rho u}\,du \\
&< 0 + \int_t^{T_s} \rho T_s\,(s - \mu_h(u))\,e^{\rho u}\,du + \int_{T_s}^{\phi(t)} \rho T_s\,(s - \mu_h(u))\,e^{\rho u}\,du \\
&< 0,
\end{aligned}$$

where the third line follows because, for $u \in [t, T_s]$, $s - \mu_h(u)$ is positive so $u\,(s - \mu_h(u))$ is bounded above by $T_s\,(s - \mu_h(u))$. For $u \in [T_s, \phi(t)]$, on the other hand, $s - \mu_h(u)$ is negative so $u\,(s - \mu_h(u))$ is also bounded above by $T_s\,(s - \mu_h(u))$. It thus follows that $\phi(t)$ is decreasing in ρ.

The order placement function is decreasing in γ. We have:

$$\frac{\partial H}{\partial \gamma} = -\int_t^{\phi} \frac{\partial \mu_h}{\partial \gamma}(u) e^{\rho u}\,du < 0,$$

because $\mu_h(z)$ increases in γ. So $\phi(t)$ decreases with γ.

A.2 Proof of Proposition 2

In this appendix we derive the dynamics of the distribution of types when investors follow the conjectured trading strategies. The analysis confirms that this results in a feasible asset allocation: at each time there is zero net trade in the market. In what follows we denote by $\mu_\sigma(t)$ the measure of investors of type σ, at time t, and we drop the time subscripts to simplify notations. The dynamics of distribution of are illustrated in Figure 5 and are summarized in the following ODEs:

$$\begin{aligned}
&\text{type } hn \quad \dot{\mu}_{hn} = -\mathrm{Mkt}_h + \mathrm{LimExec}_h + \gamma \mu_{\ell n} && (26)\\
&\text{type } ho \quad \dot{\mu}_{ho} = \mathrm{Mkt}_h + \rho \mu_{hb} + \gamma \mu_{\ell o} && (27)\\
&\text{type } hb \quad \dot{\mu}_{hb} = -\rho \mu_{hb} - \mathrm{LimExec}_h + \gamma \mu_{\ell b} && (28)\\
&\text{type } \ell n \quad \dot{\mu}_{\ell n} = \mathrm{Mkt}_\ell + \mathrm{LimExec}_\ell - \gamma \mu_{\ell n} && (29)\\
&\text{type } \ell o \quad \dot{\mu}_{\ell o} = -\rho \mu_{\ell o} - \gamma \mu_{\ell o} && (30)\\
&\text{type } \ell b \quad \dot{\mu}_{\ell b} = -\rho \mu_{\ell b} - \mathrm{LimExec}_\ell + \mathrm{LimSub} - \gamma \mu_{\ell b}, && (31)
\end{aligned}$$

where

- Mkt_h is the flow of market buy orders submitted by hn investors who contact the market.
- Mkt_ℓ is the flow of market sell orders submitted by either ℓo or ℓb investors who contact the market.
- LimSub is the flow of new limit orders submitted by either ℓo or ℓb investors who contact the market.
- LimExec_ℓ (LimExec_h) are the flow of limit sell orders executed from the book, held by low (high) valuation investors.

For instance, on the right-hand side of equation (26), the first term is the flow of hn investors who buy one unit of the asset with a market order, making a transition to the ho type. The second term is the flow of hb investors who see their limit-sell order executed, and make a transition to the hn type. The ODEs reflect features of investors' trading strategies: hn investors place market buy orders, ho investors stay put, hb investors cancel their limit orders, ℓn investors stay put. Also, ℓo and ℓb investors either place market or limit sell orders, implying that:

$$\text{LimSub} + \text{Mkt}_\ell = \rho\left(\mu_{\ell o} + \mu_{\ell b}\right). \tag{32}$$

The market clearing condition is that $\mu_{ho} + \mu_{hb} + \mu_{\ell o} + \mu_{\ell b} = s$ at all times. Taking derivatives, using the ODEs (27), (28), (30) and (31), we obtain the natural condition:

$$\begin{aligned} \text{Mkt}_h &= \text{Mkt}_\ell + \rho\mu_{\ell o} + \rho\mu_{\ell b} - \text{LimSub} \\ &= \text{Mkt}_\ell + \text{LimExec}_\ell + \text{LimExec}_h \end{aligned} \tag{33}$$

after plugging in equation (32). That is, the flow of market buy orders has to be equal to the flow of market sell orders, plus the flow of limit sell orders executed from the book. We proceed by an analysis of the three time intervals, $[0, T_s]$, $[T_s, T_f]$, and $[T_f, \infty)$.

Interval $[0, T_s]$. All hn investors buy one unit of the asset, so $\text{Mkt}_h = \rho\mu_{hn}$. In addition, limit orders are not executed so $\text{LimExec}_\ell = \text{LimExec}_h = 0$. Plugging this in the market clearing condition (33), we obtain that $\text{Mkt}_\ell = \rho\mu_{hn}$. Next, plugging in (32), we obtain that

$$\begin{aligned} \text{LimSub} &= \rho\mu_{\ell o} + \rho\mu_{\ell b} - \rho\mu_{hn} \\ &= \rho\left(\mu_{\ell o} + \mu_{\ell b} + \mu_{ho} + \mu_{hb} - \mu_{ho} - \mu_{hb} - \mu_{hn}\right) \\ &= \rho\left(s - \mu_h\right) \geq 0 \end{aligned}$$

because $t \leq T_s$. This confirms the formula we obtained in the text for the flow of limit orders submitted during $[t, t + dt]$.

Interval $[T_s, T_f]$. All hn investors who contact the market submit market buy orders, so $\text{Mkt}_h = \rho\mu_{hn}$. All ℓo and ℓb investors who contact the market submit market sell orders, so $\text{LimSub} = 0$ and $\text{Mkt}_\ell = \rho\mu_{\ell o} + \rho\mu_{\ell b}$. It thus follows from the market clearing condition (33) that:

$$\begin{aligned} \text{LimExec}_h + \text{LimExec}_\ell &= \rho\mu_{hn} - \rho\mu_{\ell o} - \rho\mu_{\ell b} \\ &= \rho\left(\rho\mu_{hn} + \mu_{ho} + \mu_{hb} - \mu_{ho} - \mu_{hb} - \mu_{\ell o} - \mu_{\ell b}\right) \\ &= \rho\left(\mu_h - s\right) \geq 0 \end{aligned}$$

because $t \geq T_s$. To figure out the values of LimExec_h and LimExec_ℓ, recall that orders executed at time t where all submitted at time $\phi^{-1}(t)$, by some low-valuation investors. Thus, the probability that an order submitted at time $\phi^{-1}(t)$ is, at time t, held by a high-valuation investor is $\pi_h(\phi^{-1}(t), t)$. By the law of large numbers, this is also the fractions of limit order executed at time t, held by high-valuation investors. To sum up:

$$\begin{aligned} \text{LimExec}_h &= \rho\,(\mu_h(t) - s)\,\pi_h(\phi^{-1}(t), t) \\ \text{LimExec}_\ell &= \rho\,(\mu_h(t) - s) - \text{LimExec}_h. \end{aligned}$$

Interval $[T_f, \infty)$. There is no activity in the limit order book so $\text{LimExec}_\ell = \text{LimExec}_h = \text{LimSub} = 0$. All low-valuation investors submit market sell orders, so $\text{Mkt}_\ell = \rho\mu_{\ell o}$. These are matched by an equal flow of market buy orders from hn investors, so $\text{Mkt}_h = \rho\mu_{\ell o}$.

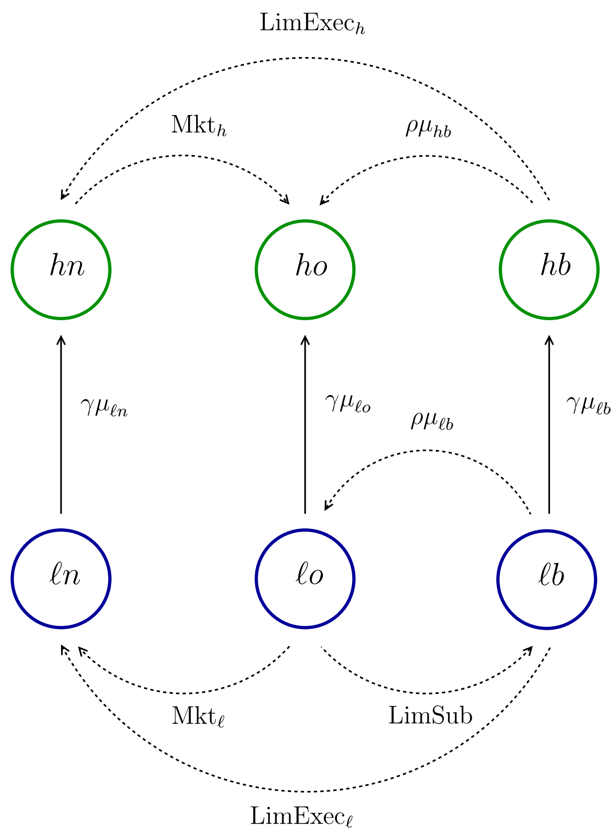

Figure 5: Inflows and outflows between types

A.3 Proof of Proposition 3

The ODE for the price in the interval $[T_s, T_f]$ is given by (7). We now turn to the ODE during the interval $[0, T_s]$. Starting from equation (11), we obtain:

$$
\begin{aligned}
0 &= -p(t) + \mathbb{E}_t\left[\int_t^{\min\{\tau,\phi(t)\}} e^{-r(u-t)}\theta(u)\,du + e^{-r(\min\{\tau,\phi(t)\}-t)}p(\min\{\tau,\phi(t)\})\right] && (34)\\
&= \mathbb{E}_t\left[\int_t^{\min\{\tau,\phi(t)\}} e^{-r(u-t)}\left(\theta(u)+\dot{p}(u)-rp(u)\right)du\right] && (35)\\
&= \mathbb{E}_t\left[\int_t^{\phi(t)} \mathbb{I}_{\{u\le\tau\}}e^{-r(u-t)}\left(\theta(u)+\dot{p}(u)-rp(u)\right)du\right] && (36)\\
&= \int_t^{\phi(t)} \mathbb{E}_t\left[\mathbb{I}_{\{u\le\tau\}}\right]e^{-r(u-t)}\left(\mathbb{E}_t\left[\theta(u)\right]+\dot{p}(u)-rp(u)\right)du && (37)\\
&= \int_t^{\phi(t)} e^{-(r+\rho)(u-t)}\left(1-\delta+\delta\pi_h(t,u)+\dot{p}(u)-rp(u)\right)du && (38)
\end{aligned}
$$

where equation (34) obtains by rearranging equation (11), equation (35) obtains because

$$p(t_2)e^{-r(t_2-t_1)} - p(t_1) = \int_{t_1}^{t_2}\left(\dot{p}(u)-rp(u)\right)e^{-r(u-t_1)},$$

equation (37) obtains because of the independence between the contact time and the switching time, and equation (38) follows from the fact that Proba$(\tau \ge u \,|\, \tau \ge t) = e^{-\rho(u-t)}$ and the definition of $\pi_h(t,u)$. Now note that, since $\phi(T_s) = T_s$, the right-hand side of equation (38) is clearly equal to zero at $t = T_s$. Thus, for equation (38) to be also satisfied at all $t \le T_s$, it must be that its derivative with respect to t is also equal to zero:

$$
\begin{aligned}
0 &= -\Big[1-\delta+\delta\pi_h(t,t)+\dot{p}(t)-rp(t)\Big]\\
&\quad +\phi'(t)\Big[1-\delta+\delta\pi_h(t,\phi(t))+\dot{p}(\phi(t))-rp(\phi(t))\Big]e^{-(r+\rho)(\phi(t)-t)} && (39)\\
&\quad +(r+\rho)\int_t^{\phi(t)} e^{-(r+\rho)(u-t)}\Big(1-\delta+\delta\pi_h(t,u)+\dot{p}(u)-rp(u)\Big)du\\
&\quad +\delta\int_t^{\phi(t)} e^{-(r+\rho)(u-t)}\frac{\partial\pi_h}{\partial t}(t,u)\,du\\
\Leftrightarrow 0 &= -\left[1-\delta+\dot{p}(t)-rp(t)\right]+0+0+\delta\int_t^{\phi(t)} e^{-(r+\rho)(u-t)}\frac{\partial\pi_h}{\partial t}(t,u)\,du\\
\Leftrightarrow rp(t) &= 1-\delta-\delta\int_t^{\phi(t)} e^{-(r+\rho)(u-t)}\frac{\partial\pi_h}{\partial t}(t,u)\,du+\dot{p}(t). && (40)
\end{aligned}
$$

Note that the second term is equal to zero because of equation (7). The third term is also equal to zero because of equation (11). Using the functional form $\pi_h(t,u) = 1 - e^{-\gamma(u-t)}$, we have that, for $t \in [0, T_s]$:

$$rp(t) = 1-\delta+\frac{\delta\gamma}{r+\rho+\gamma}\left(1-e^{-(r+\rho+\gamma)(\phi(t)-t)}\right)+\dot{p}(t)$$

Puting the different results together the price function is as follows:

$$
\begin{aligned}
t \in [0, T_s] \qquad & rp(t) = 1 - \delta + \frac{\delta\gamma}{r+\rho+\gamma}\left(1 - e^{-(r+\rho+\gamma)(\phi(t)-t)}\right) + \dot{p}(t) \\
t \in [T_s, T_f] \qquad & rp(t) = 1 - \delta + \delta\left(1 - e^{-\gamma\left(t-\phi^{-1}(t)\right)}\right) + \dot{p}(t) \\
t \geq T_f \qquad & rp(t) = 1.
\end{aligned}
$$

Note also that the first two ODEs imply that, at $t = T_s$:

$$\dot{p}(T_s) = rp(T_s) - (1-\delta),$$

because $\phi(T_s) = T_s$. So the price process is continuously differentiable at $t = T_s$. Also, the last two equations imply that, at $t = T_f$:

$$\dot{p}(T_f^-) = \delta\left(1 - e^{-\gamma T_f}\right) > 0,$$

because $p(T_f) = 1/r$ and $\phi^{-1}(T_f) = 0$ Note that, since $\phi(t)$ is continuously differentiable, the ODE is continuously differentiable on each interval, implying that the solution, $p(t)$ is twice continuously differentiable on each interval. Thus, the derivative of the price $d(t) = \dot{p}(t)$ solves the ODEs:

$$
\begin{aligned}
t \in [0, T_s] \qquad & rd(t) = \delta\gamma\Big(\phi'(t) - 1\Big)e^{-(r+\rho+\gamma)(\phi(t)-t)} + \dot{d}(t) \\
t \in [T_s, T_f] \qquad & rd(t) = \delta\gamma\left(1 - \frac{1}{\phi' \circ \phi^{-1}(t)}\right)e^{-\gamma\left(t-\phi^{-1}(t)\right)} + \dot{d}(t).
\end{aligned}
$$

Let us start with the interval $[T_s, T_f]$. Integrating the ODE for $d(t)$, we obtain that:

$$d(t) = \int_t^{T_f} \delta\gamma\left(1 - \frac{1}{\phi' \circ \phi^{-1}(u)}\right)e^{-\gamma\left(u-\phi^{-1}(u)\right)}e^{-r(u-t)}\,du + e^{-r(T_f-t)}\dot{p}(T_f^-) > 0$$

which is positive because, since $\phi^{-1}(u)$ is strictly decreasing, the integrand is positive, and $\dot{p}(T_f^-) > 0$. Next, consider the initial time interval, $[0, T_s]$. After integrating the ODE for $d(t)$, we obtain:

$$
\begin{aligned}
d(t) \quad = \quad & \delta\gamma\int_t^{T_s} e^{-r(u-t)}\Big(\phi'(u) - 1\Big)e^{-(r+\rho+\gamma)(\phi(u)-u)}\,du \\
& +\delta\gamma\int_{T_s}^{\phi(t)} e^{-r(u-t)}\left(1 - \frac{1}{\phi' \circ \phi^{-1}(u)}\right)e^{-\gamma\left(u-\phi^{-1}(u)\right)}\,du \\
& +e^{-r(\phi(t)-t)}d(\phi(t)).
\end{aligned}
$$

Now we make the change of variable $v = \phi^{-1}(u)$ in the second integral. We obtain:

$$
\begin{aligned}
d(t) \quad = \quad & \delta\gamma\int_t^{T_s} e^{-r(u-t)}\Big(\phi'(u) - 1\Big)e^{-(r+\rho+\gamma)(\phi(u)-u)}\,du \\
& +\delta\gamma\int_t^{T_s} e^{-r(\phi(v)-t)}\Big(1 - \phi'(v)\Big)e^{-\gamma(\phi(v)-v)}\,dv \\
& +e^{-r(\phi(t)-t)}d(\phi(t)).
\end{aligned}
$$

After collecting the first two lines, we obtain:

$$d(t) \quad = \quad \delta\gamma \int_t^{T_s} e^{-r(\phi(u)-t)-\gamma(\phi(u)-u)} \Big(1-\phi'(u)\Big) \left(1-e^{-\rho(\phi(u)-u)}\right) \, du + e^{-r(\phi(t)-t)} d(\phi(t)).$$

The integrand in the first term is positive because $\phi'(u) < 0$ and $\phi(u) \geq u$. We also have $d(\phi(t)) \geq 0$ since $\phi(t) \geq T_s$. So $d(t) > 0$, meaning that the price is indeed increasing.

A.4 Proof of Lemma 2

Consider first investors of types ℓo and ℓb. Their prescribed strategy is to sell the asset, with a continuation utility $V_{\ell n}(t) + p(t)$. During the initial interval $[0, T_s]$, another optimal strategy is to submit a limit order to sell at time $\phi(t)$. The possible deviation is to submit a limit order to sell at some time $z \geq t$, with a continuation utility $V_{\ell b}(t, z)$.

Similarly, for an investor of type ℓn, the prescribed strategy is to stay put, with a continuation utility $V_{\ell n}(t)$, and the possible deviation is to buy the asset and immediately submit a limit order to sell at time $z \geq t$, with a continuation utility $V_{\ell b}(t, z)$.

In both cases, one sees that the prescribed strategies are optimal if

$$V_{\ell b}(t, z) - V_{\ell n}(t) - p(t) \leq 0.$$

Moreover, since $V_{\ell b}(t, t) = V_{\ell n}(t) + p(t)$, the upper bound is achieved at $z = t$. Note that submitting a limit order to sell at time $\phi(t)$ is also optimal if the upper bound of zero is achieved at $z = \phi(t)$.

A.5 Proof of Lemma 3

Consider first an investor of type hn. His prescribed strategy is to buy the asset, with a continuation utility $V_{ho}(t) - p(t)$. The possible deviation is to buy but immediately submit a limit order to sell at time $z \geq t$, with continuation utility $V_{hb}(t, z) - p(t)$. (Note that staying put corresponds to $z = t$).

Next consider an investor of type ho or hb. The prescribed strategy is to cancel any outstanding limit order and stay put, with a continuation utility $V_{ho}(t)$. The possible deviation is to submit a limit order to sell at time $z \geq t$, with continuation utility $V_{hb}(t, z)$.

In both cases, one sees that the prescribed strategy is optimal if

$$V_{hb}(t, z) - V_{ho}(t) \leq 0.$$

Note also that $V_{hb}(t, z) \to V_{ho}(t)$, as $z \to \infty$.

A.6 Proof of Proposition 4

To prove the optimality of trading strategies, we first calculate the continuation utilities, and we proceed with an application of the optimality verification lemmas, 2 and 3

A.6.1 Low-valuation investor

Net utility. To apply the optimality verification Lemma 2, we need to calculate the net utility $V_{\ell b}(t,z) - V_{\ell o}(t) - p(t)$, for any $z \geq t$. We start with

$$
\begin{aligned}
V_{\ell b}(t,z) = \mathbb{E}_t\Bigg[& \int_t^{\min\{z,\tau\}} e^{-r(u-t)}\theta(u)\,du \\
& + \mathbb{I}_{\{z\leq\tau \text{ and } \theta(z)=h\}}\; e^{-r(z-t)}\left(p(z)+V_{hn}(z)\right) \\
& + \mathbb{I}_{\{z\leq\tau \text{ and } \theta(z)=\ell\}}\; e^{-r(z-t)}\left(p(z)+V_{\ell n}(z)\right) \\
& + \mathbb{I}_{\{\tau\leq z \text{ and } \theta(z)=h\}}\; e^{-r(\tau-t)}V_{ho}(\tau) \\
& + \mathbb{I}_{\{\tau\leq z \text{ and } \theta(z)=\ell\}}\; e^{-r(\tau-t)}\Big(V_{\ell n}(\tau)+p(\tau)\Big) \Bigg],
\end{aligned}
$$

where τ denotes the next contact time with the market. With a slight abuse of notation, "$\theta(u) = \ell$" means $\theta(u) = 1-\delta$, and "$\theta(u) = h$" means $\theta(u) = 1$. The first term is the expected present value of flow utilities that a ℓb investor enjoys until $\min\{z,\tau\}$. The other terms are the continuation utilities, given the conjectured trading strategies. There are four possible continuation utilities, depending on whether the execution time, z, is attained before or after the next contact time τ with the market ($\min\{z,\tau\} = z$ or $= \tau$), and on whether the investor has a high or low utility at $\min\{z,\tau\}$. Similarly, for any $z \geq t$, we have

$$
\begin{aligned}
V_{\ell n}(t) = \mathbb{E}_t\Bigg[& \mathbb{I}_{\{z\leq\tau \text{ and } \theta(z)=h\}}\; e^{-r(z-t)}V_{hn}(z) \\
& + \mathbb{I}_{\{z\leq\tau \text{ and } \theta(z)=\ell\}}\; e^{-r(z-t)}V_{\ell n}(z) \\
& + \mathbb{I}_{\{\tau\leq z \text{ and } \theta(z)=h\}}\; e^{-r(\tau-t)}\Big(V_{ho}(\tau)-p(z)\Big) \\
& + \mathbb{I}_{\{\tau\leq z \text{ and } \theta(z)=\ell\}}\; e^{-r(\tau-t)}V_{\ell n}(\tau)\Bigg].
\end{aligned}
$$

Taking the difference between the two, $V_{\ell b}(t,z) - V_{\ell n}(t)$, and subtracting the price, we obtain:

$$
V_{\ell b}(t,z) - V_{\ell n}(t) - p(t) = -p(t) + \mathbb{E}_t\left[\int_t^{\min\{z,\tau\}} e^{-r(u-t)}\theta(u)\,du + e^{-r(\min\{z,\tau\}-t)}p(\min\{z,\tau\})\right].
$$

The expression can be simplified further as follows:

$$
\begin{aligned}
& V_{\ell b}(t,z) - V_{\ell n}(t) - p(t) \\
= \;& \mathbb{E}_t\left[\int_t^{\min\{z,\tau\}} e^{-r(u-t)}\Big(\theta(u)+\dot{p}(u)-rp(u)\Big)\,du\right] && (41) \\
= \;& \int_t^z e^{-(r+\rho)(u-t)}\left(1-\delta+\delta\pi_h(t,u)+\dot{p}(u)-rp(u)\right)\,du, && (42)
\end{aligned}
$$

by following the exact same steps as for equation (40), in the proof of Proposition 3.

The marginal value of execution time, $z \in [T_s, T_f]$**.** We proceed with a calculation of the marginal value of increasing the execution time for $z \in [T_s, \infty)$. Taking the derivative of (42) with respect to z, we obtain:

$$
\begin{aligned}
& e^{-(r+\rho)(z-t)}\Big(1-\delta+\delta\pi_h(t,z)+\dot{p}(z)-rp(z)\Big) && (43)\\
= \quad & \delta e^{-(r+\rho)(z-t)}\Big(\pi_h(t,z)-\pi_h(\phi^{-1}(z),z)\Big)\\
= \quad & \delta e^{-(r+\rho)(z-t)}\left(\frac{\mu_h(z)-\mu_h(t)}{1-\mu_h(t)}-\frac{\mu_h(z)-\mu_h(\phi^{-1}(z))}{1-\mu_h(\phi^{-1}(z))}\right)\\
= \quad & \delta e^{-(r+\rho)(z-t)}\frac{1-\mu_h(z)}{(1-\mu_h(t))(1-\mu_h(\phi^{-1}(z)))}\Big(\mu_h(\phi^{-1}(z))-\mu_h(t)\Big), && (44)
\end{aligned}
$$

where the second line follows from substituting in the ODE (7) for the price, $z \in [T_s, T_f]$, and the last line from simple algebraic manipulations.

The marginal value of execution time, $z > T_f$. Using equation (43), one sees that the marginal value of increasing the execution time at $z > T_f$ is

$$
e^{-(r+\rho)(z-t)}\Big(-1+\pi_h(t,z)\Big) < 0, \tag{45}
$$

since $p(z) = 1/r$.

Optimality when $t > T_s$. For all $t > T_s$, one sees from equation (44) that, if $z \in [t, T_f]$, then $\phi^{-1}(z) \le T_s < t$, and since $\mu_h(t)$ is increasing, it follows that the marginal value of increasing execution time is negative. From equation (45), it is clear that the marginal value is negative at $z \ge T_f$ as well. Taken together, it follows that, for $t > T_s$, the best sell order is a marketable limit order, i.e. it is best to pick the execution time $z = t$.

Optimality when $t \le T_s$. From equation (43), it follows that the marginal value of execution time is positive for $z \in [T_s, \phi(t))$, zero at $z = \phi(t)$, and negative for $z \in (\phi(t), T_f]$. Thus, the best execution time in $[T_s, T_f]$ is $z = \phi(t)$. Moreover, comparing (41) and (35), one sees that, by construction of the price path, low-valuation investors are indifferent:

$$
V_{\ell b}(t,\phi(t)) - V_{\ell n}(t) - p(t) = 0.
$$

That is, the utility $V_{\ell b}(t,\phi(t))$ at time t of submitting a limit sell order to be executed at time $\phi(t)$ is exactly equal to the utility $V_{\ell n}(t)$ of submitting a marketable sell order at time t. The only thing that remains to be shown is that $V_{\ell b}(t,z) - V_{\ell n}(t) - p(t) \le 0$ for all $z \in [t, T_s]$. We proceed as follows. We fix some $z \in [t, T_s]$ and we plug (40) into equation (42):

$$
V_{\ell b}(t,z) - V_{\ell n}(t) - p(t) \quad = \quad \delta \int_t^z e^{-(r+\rho)(u-t)}\left(\pi_h(t,u) + \int_u^{\phi(u)} e^{-(r+\rho)(v-u)}\frac{\partial \pi_h}{\partial u}(u,v)\,dv\right) du.
$$

Keeping in mind that $\pi_h(v,v)=0$, the first integral term can be written:

$$\delta\int_t^z e^{-(r+\rho)(v-t)}\pi_h(t,v)\,dv = \delta\int_t^z e^{-(r+\rho)(v-t)}\left(-\int_t^v \frac{\partial \pi_h}{\partial u}(u,v)\,du\right)dv$$
$$= \quad -\delta\int_t^z\int_u^z e^{-(r+\rho)(v-t)}\frac{\partial \pi_h}{\partial u}(u,v)\,dv\,du, \tag{46}$$

where the third equality follows from exchanging the order of integration. The second integral term, on the other hand, is:

$$\delta\int_t^z e^{-(r+\rho)(u-t)}\int_u^{\phi(u)} e^{-(r+\rho)(v-u)}\frac{\partial \pi_h}{\partial u}(u,v)\,dv\,du$$
$$= \quad \delta\int_t^z\int_u^{\phi(u)} e^{-(r+\rho)(v-t)}\frac{\partial \pi_h}{\partial u}(u,v)\,dv\,du \tag{47}$$

Adding up the two integrals above, and keeping in mind that $z \le T_s \le \phi(u)$, we obtain

$$V_{\ell b}(t,z) - V_{\ell n}(t) - p(t) = \delta\int_t^z\int_z^{\phi(u)} e^{-(r+\rho)(v-t)}\frac{\partial \pi_h}{\partial u}(u,v)\,dv\,du$$

which is negative because $\pi_h(u,v)$ is decreasing in its first argument.

A.6.2 High-valuation investors

The continuation utility of a high-valuation non-owner, hn, is

$$V_{hn}(t) = \mathbb{E}_t\left[e^{-r(\tau-t)}\left(\frac{1}{r}-p(\tau)\right)\right], \tag{48}$$

since a hn investor buys at his first contact time with the market. The continuation utility of a high-valuation owner is $V_{ho}(t)=1/r$. Thus,

$$V_{ho}(t) - V_{hn}(t) - p(t) = \mathbb{E}_t\left[\int_t^\tau\left(1+\dot{p}(u)-rp(u)\right)\right] \ge 0, \tag{49}$$

because $\dot{p}(u)\ge 0$ and $p(u)\le 1/r$. Now

$$V_{hb}(t,z) - V_{ho} = \mathbb{E}_t\left[e^{-r(z-t)}\mathbb{I}_{\{z\le\tau\}}\left(V_{hn}(z)+p(z)-V_{ho}\right)\right]. \tag{50}$$

Indeed, the only scenario in which the continuation utilities differ is when $z\le\tau$: if the investor has a limit order, his continuation utility is $V_{hn}(z)+p(z)$. If he does not, his continuation utility is $V_{ho}(z)$. Clearly, because of equation (49), this is negative. It also goes to zero as z goes to infinity.

A.7 Proof of Proposition 5

This follows directly from applying the result of Appendix B.1 to equation (21).

A.8 Proof of Proposition 6

The only thing that remains to be shown that, for ρ large enough, the number of cancelation increases with ρ. To see this first note that the derivative of the flow of cancelations with respect to ρ is:

$$\frac{\partial C}{\partial \rho} = L(t) + \rho \frac{\partial L(t)}{\partial \rho}.$$

Also note that

$$\frac{\partial L}{\partial \rho} = \int_0^t (1 + \rho(u-t))\, e^{\rho(u-t)}\, (s - \mu_h(u))\; du.$$

Thus, an application of the result of Appendix B.2 implies that $\rho^2 \partial L/\partial \rho$ goes to $-\dot{\mu}_h(t)$,as ρ goes to infinity. Hence $\partial L \partial \rho$ goes to 0 and, for ρ large enough, $\partial C/\partial \rho$ goes to $s - \mu_h(t)$ which is positive for $t < T_s$.

A.9 Proof of Proposition 7

We show that the ratio $M(t)/V(t)$ increases with ρ, as long as ρ is large enough. This is equivalent to show that

$$\frac{d}{d\rho} \log \Big[\rho(s - \mu_h(t) + L(t)) \Big] > \frac{d}{d\rho} \log V(t),$$

for ρ large enough. The left hand side of the above expression is

$$\begin{aligned} \frac{s - \mu_h(t) + L(t) + \rho \frac{\partial L}{\partial \rho}}{\rho \Big(s - \mu_h(t) + L(t) \Big)} &= \frac{1}{\rho} \frac{s - \mu_h(t) + L(t) + o(1)}{s - \mu_h(t) + L(t)} \\ &= \frac{1}{\rho} + o\left(\frac{1}{\rho}\right). \end{aligned} \tag{51}$$

Now using equation (20), we obtain that

$$\begin{aligned} \frac{1}{V(t)} \frac{\partial V}{\partial \rho} &= \frac{1}{\rho} - \frac{1}{\rho + \gamma} + \frac{t e^{-(\rho+\gamma)t}}{1 - e^{-(\rho+\gamma)t}} \\ &= \frac{\gamma}{\rho(\rho+\gamma)} + \frac{t e^{-(\rho+\gamma)t}}{1 - e^{-(\rho+\gamma)t}} \\ &= o\left(\frac{1}{\rho}\right). \end{aligned} \tag{52}$$

Comparing (51) and (52), it the follows that $M(t)/V(t)$ increases for ρ large enough.

B Two useful results

B.1 First result

The statement Let $f(z)$ be a bounded and integrable function which is continuous at some $t \geq 0$. Then

$$\int_0^t \rho f(u) e^{\rho(u-t)} \to f(t)$$

as ρ goes to infinity.

The proof. Pick η such that $|f(u) - f(t)| < \varepsilon$ for all $t - \eta \leq u \leq t$.

$$
\begin{aligned}
& \left| \int_0^t \rho e^{\rho(u-t)} f(u)\, du - f(t) \right| \\
= \quad & \left| \int_0^t \rho e^{\rho(u-t)} \left(f(u) - f(t)\right)\, du + f(t) \left(\int_0^t \rho e^{\rho(u-t)}\, du - 1 \right) \right| \\
\leq \quad & \left| \int_0^{t-\eta} \rho e^{\rho(u-t)} \left(f(u) - f(t)\right)\, du + \int_{t-\eta}^t \rho e^{\rho(u-t)} |f(u) - f(t)|\, du - f(t) e^{-\rho t} \right| \\
\leq \quad & 2 \sup |f(u)| \int_0^{t-\eta} \rho e^{\rho(u-t)}\, du + \varepsilon \int_{t-\eta}^t \rho e^{\rho(u-t)}\, du - f(t) e^{-\rho t} \\
\leq \quad & 2 \sup |f(u)| \left(e^{-\rho\eta} - e^{-\rho t} \right) + \varepsilon - f(t) e^{-\rho t} \\
\leq \quad & 2\varepsilon,
\end{aligned}
$$

for ρ large enough.

B.2 Second useful result

The statement. Suppose that $f(u)$ is twice continuously differentiable over $[0, t]$. Then

$$
\rho^2 \int_0^t f(u) \left[1 + \rho(u-t)\right] e^{\rho(u-t)}\, du \to f'(t),
$$

as ρ goes to infinity.

The proof. We start with a first integration by part, noting that:

$$
\frac{d}{du} (u-t) e^{\rho(u-t)} = \left[1 + \rho(u-t)\right] e^{\rho(u-t)}.
$$

This shows that

$$
\begin{aligned}
& \rho^2 \int_0^t f(u) \left[1 + \rho(u-t)\right] e^{\rho(u-t)}\, du = \rho^2 \left[f(u)(u-t) e^{\rho(u-t)} \right]_0^t - \rho^2 \int_0^t f'(u)(u-t) e^{\rho(u-t)}\, du \\
= \quad & \rho^2 f(0) t e^{-\rho t} - \rho^2 \int_0^t (u-t) f'(u) e^{\rho(u-t)}\, du.
\end{aligned}
$$

We integrate the second term by part again, differentiating $\rho(u-t)f'(u)$ and integrating $\rho e^{\rho(u-t)}$. We obtain:

$$
\begin{aligned}
& \rho^2 \int_0^t f(u) \left[1 + \rho(u-t)\right] e^{\rho(u-t)}\, du \\
= \quad & \rho^2 f(0) t e^{-\rho t} - \rho \left[(u-t) f'(u) e^{\rho(u-t)} \right]_0^t + \int_0^t \left[f''(u)(u-t) + f'(u)\right] \rho e^{\rho(u-t)}\, du \\
= \quad & \rho^2 f(0) t e^{-\rho t} + \rho f'(0) t e^{-\rho t} + \int_0^t \left[f''(u)(u-t) + f'(u)\right] \rho e^{\rho(u-t)}\, du.
\end{aligned}
$$

The result then follows by noting that the first two terms go to zero as ρ goes to infinity, and by applying the result of section B.1 to the third term.

References

Antje Berndt, Rohan Douglas, Darrell Duffie, and David Ferguson, Mark andSchranz. Measuring default risk premia from default swap rates and edfs. 2005. 2

Bruno Biais, Piere Hillion, and Chester Spatt. An empirical analysis of the limit order book and the order flow in the paris bourse. *Journal of Finance*, 50:1655–1689, 1995. 4, 25

Joshua D. Coval and Erik Stafford. Asset fire sales (and purchases) in equity markets. *Journal of Financial Economics*, 86:479–512, 2007. 2

Zhi Da and Pengjie Gao. Clientele change, liquidity shock, and the return on financially distressed stocks. Working Paper, Northwestern University, 2007. 4

Harold Demsetz. The cost of transacting. *The Quarterly Journal of Economics*, 82:33–53, 1968. 2

Darrell Duffie, Nicolae Gârleanu, and Lasse H. Pedersen. Securities lending, shorting, and pricing. *Journal of Financial Economics*, 66:307–339, 2002. 5

Darrell Duffie, Nicolae Gârleanu, and Lasse H. Pedersen. Over-the-counter markets. *Econometrica*, 2005. 3, 5, 11

Darrell Duffie, Nicolae Gârleanu, and Lasse H. Pedersen. Valuation in over-the-counter markets. *Review of Financial Studies*, 20:1865–1900, 2007. 3, 5

Andrew Ellul, Graig W. Holden, Pankaj K. Jain, and Robert H. Jennings. Order dynamics: Recent evidence from the nyse. Working paper, Indiana University Bloomington, 2007. 25

Thierry Foucault. Order flow composition and trading costs in a dynamic limit order market. *Journal of Financial Markets*, 2:99–134, 1999. 5

Thierry Foucault, Ohad Kadan, and Eugene Kandel. Limit order book as a market for liquidity. *Review of Financial Studies*, 18:1171–1217, 2005. 5

Kenneth D. Garbade and William L. Silber. Structural organization of secondary markets: Clearing frequency, dealer activity, and liquidity risk. *Journal of Finance*, 34:577–593, 1979. 2

Nicolae Gârleanu. Portfolio choice and pricing in illiquid markets. 2008. 5

Robin Greenwood. Short and long-term demand curves for stocks: Theory and evidence on the dynamics of arbitrage. *Journal of Financial Economics*, 75:607–650, 2005. 2

Mark D. Griffiths, Brian F. Smith, D. Alasdair Turnbull, and Robert W. White. The costs and determinants of order aggressiveness. *Journal of Financial Economics*, 56:65–88, 2000. 25

Sanford J. Grossman and Merton H. Miller. Liquidity and market structure. *Journal of Finance*, 43:617–637, 1988. 2

Larry Harris. *Trading and Exchanges: Market Microstructure for Practitioners*. Oxford University Press, New York, New York, 2003. 2, 6

Joel Hasbrouck and Gideon Saar. Limit orders and volatility in a hybrid market: The island ecn. Working paper, NYU, 2002. 5

Terrence Hendershott, Charles M. Jones, and Albert J. Menkveld. Does algorithmic trading improve liquidity? Working paper, UC Berkeley, 2007. 5, 27, 28

Amir E. Khandaniy and Andrew W. Lo. What happened to the quants in August 2007? Working paper, MIT, 2008. 2, 4

Ricardo Lagos and Guillaume Rocheteau. Liquidity in asset markets with search frictions. *Econometrica*, 2009. Forthcoming. 5

Gara Mínguez Afonso. Liquidity and congestion. 2008. 5

Christine A. Parlour. Price dynamics in limit order markets. *Review of Financial Studies*, 11:789–816, 1998. 5

Christine A. Parlour and Duane J. Seppi. Limit order markets: A survey. In A.W.A. Boot and A.V. Thakor, editors, *Handbook of Financial Intermediation and Banking,*. 2008. 5

Iaonid Rosu. A dynamic model of the limit-order book. *Review of Financial Studies*, 2009. Forthcoming. 5, 6

Yeneng Sun. The exact law of large numbers via fubini extension and characterization of insurable risks. *Journal of Economic Theory*, 126:31–69, 2006. 6

Dimitri Vayanos and Tan Wang. Search and endogenous concentration of liquidity in asset markets, working paper. *Journal of Economic Theory*, 136:66–104, 2007. 5

Dimitri Vayanos and Pierre-Olivier Weill. A search-based theory of the on-the-run phenomenon. *Journal of Finance*, 63:1361–1398, 2008. 5

Pierre-Olivier Weill. Leaning against the wind. *Review of Economic Studies*, 74:1329–1354, 2007. 3, 5

Pierre-Olivier Weill. Liquidity premia in dynamic bargaining markets. *Journal of Economic Theory*, 140:66–96, 2008. 5