Liquidity shocks and order book dynamics*

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

May 20, 2009

Abstract

We propose a dynamic competitive equilibrium model of limit order trading, based on the premise that investors cannot monitor markets continuously. We study how limit order markets absorb transient liquidity shocks, which occur when a significant fraction of investors lose their willingness and ability to hold assets. We characterize the equilibrium dynamics of market prices, bid-ask spreads, order submissions and cancelations, as well as the volume and limit order book depth they generate.

Keywords: Limit-order book, liquidity, bid-ask spread, search
JEL Codes: G12, D83

*Many thanks, for helpful discussions and suggestions, to Andy Atkeson, Dirk Bergemann, Darrell Duffie, Emmanuel Farhi, Thierry Foucault, Christian Hellwig, Hugo Hopenhayn, Vivien Lévy–Garboua, Johannes Horner, John Moore, Henri Pages, Jean Charles Rochet, Larry Samuelson, Jean Tirole, Aleh Tsyvinski, Jusso Valimaki, Dimitri Vayanos, Adrien Verdelhan, and Glen Weyl; and seminar participants at the Federal Reserve Bank of Minneapolis, UCLA Economics, Toulouse, LSE, LBS, the Banque de France Conference on Macroeconomics and Liquidity, UCSD Rady School of Business, the University of Minnesota Carlson School of Business, UCLA Anderson, Yale University theory and macro, Harvard University, Loyola Marymount University, the FBF-IDEI Conference on Investment Banking and Financial Markets, and Boston University. This research benefited from the support of the "Financial Markets and Investment Banking Value Chain Chair" sponsored by the Fédération Bancaire Française.

[†]Toulouse School of Economics (CNRS, IDEI), biais@cict.fr.
[‡]University of California Los Angeles and NBER, poweill@econ.ucla.edu.

Investors do not trade each and every asset continuously. They face a multiplicity of tasks to which they must allocate time and effort. They must participate in meetings, interact with customers and comply with reporting requirements. In order to make efficient financial decisions, they need to collect and process information regarding asset supply and demand, as well as the fundamentals of underlying cash flows. In doing so they must decide which assets to monitor in priority. Because these activities take time, they prevent investors from continuously making trading decisions.[1]

That all investors are not continuously trading reduces liquidity (Demsetz, 1968, Garbade and Silber, 1979). However, 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 investor population is affected by a change in its willingness and ability to hold assets, as in Grossman and Miller (1988). This can occur because of changes in the characteristics of assets. For example, many institutions are required to sell bonds which lose their investment grade status, or to sell 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; for example, Funds experiencing large outflows must sell their holdings, as documented by Coval and Stafford (2007). For regulatory reasons, after large losses, banks 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 execute 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 markets to liquidity shocks, we address the following questions: How do prices react and adjust? What are the dynamics of liquidity supply and demand and the corresponding evolution of the order book, trading volume and transactions costs? What are the optimal orders for investors?

[1]Corwin and Coughenour (2008) find empirically that specialists allocate effort towards certain stocks, which results in them trading less frequently the other stocks they are entrusted with.

We study these issues in a dynamic rational expectations model: anticipating the dynamics of prices and of the order book, agents design their optimal orders. In equilibrium, these orders 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 utility flow 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 (2007) and Weill (2007) and assume that at time 0 all investors switch to the low-valuation type. Then, as time passes, 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 in this process. Efficiency would require that low-valuation investors sell to high-valuation investors. Such efficient reallocation of the asset is delayed, however, because all investors are not continuously trading on the market. To model discontinuous market presence, we follow Duffie, Gârleanu, and Pedersen (2005) and assume that final investors make contact with the market at Poisson arrival times. The greater the intensity of this Poisson process, the greater the frequency with which investors make contact with the market.

When making contact with the market, investors can place limit orders to sell or buy, and, if they already have orders in the order book, they can cancel or modify them. Marketable limit orders (i.e., sell orders at prices lower than or equal to the best bid and buy orders at prices greater than or equal to the ask) hit the market quotes and are immediately executed. Non–immediately executed limit orders are stored in the order book. The dynamics of the order book, in particular the evolution of the bid–ask spread and its 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, 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 the progressive price recovery reflect the delayed and gradual adjustment of the market due to discontinuous market presence.

High valuation investors, when making contact with 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 into two phases. In the first phase, buy orders are placed at very low prices; these set the bid quote and are hit by market orders to sell. But, as time passes buy orders 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 market 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 depth on the ask side of the order book grows, beginning at high prices, then at progressively lower prices. 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 must 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 execution time. For early investors, the probability of switching to high valuation prior to this time is higher than for investors arriving on the market later. 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 progressively lower prices.

Our theoretical analysis generates several empirical implications in line with stylized facts. Order placement activity concentrates at the best quotes and orders of similar type 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 a liquidity shock there is a sharp decline in price and strong order flow imbalance reflecting selling pressure, and then the price gradually recovers.

Progress in communications technology and the computerization of exchanges and trading rooms have reduced the cost of market access. An important recent development is algorithmic trading, which increases the speed with which investors can process information and which reduces the cost of implementing trading strategies. In the language of our model, this corresponds to an increase in the rate at which investors make contact with 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 as the market becomes more computerized. In the context of our model, this ratio can be analyzed, along with its link to the rate at which traders make contact with the market. Consistent with the evidence of Hendershott, Jones, and Menkveld, we show that as the rate at which investors make contact with the market grows larger, 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 make contact with the market grows larger, 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, cancelations have become very frequent (see Hasbrouck and Saar, 2009). Hasbrouck and Saar provide evidence that quickly canceled limit orders often correspond to modifications, whereby the order is resubmitted at another price or transformed into a market order. These two phenomena become prevalent in our equilibrium as the intensity of contact with the market grows large.

This paper is the first to introduce limit orders into the liquidity paradigm initiated by Duffie, Gârleanu, and Pedersen (2005).[2] By doing so, we add to the rich literature

[2]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 (2009) demonstrate that relaxing the asset holding constraint can have important impact on asset prices and on the entry decisions of intermediaries. Mínguez Afonso (2008) studies the externalities involved in investors' entry decisions. Weill (2007), Lagos, Rocheteau, and Weill (2007) and Weill (2008) study the liquidity provision of intermediaries in response to liquidity shocks similar to the one studied in the present model. Finally, Duffie and Manso (2007) and Golosov, Lorenzoni, and Tsyvinski (2008) study the diffusion of private information about asset payoffs in search-and-matching financial

on limit order markets (see the insightful survey by Parlour and Seppi, 2008). The first dynamic models of limit order books were offered by Foucault (1999) and Parlour (1998). The former proposes an elegant rational expectations model in which orders reflect the anticipations of traders about future market prices, but traders and orders survive only one period. The latter provides 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. Rosu (2009) presents a fully dynamic model, where traders arrive on the market at Poisson times; he provides an insightful analysis of the game played by strategic traders, and characterizes Markov perfect equilibria, where only the number of buyers and sellers matters. Because traders are large relative to the market, their orders have an impact on prices, which Rosu (2009) characterizes. While these papers take a game theoretic approach to analyze interactions between traders with market power, we study a continuum of competitive agents, where orders reflect investors' marginal valuation of the asset rather than strategic considerations. This different approach demonstrates that, with imperfect monitoring and discontinuous market presence, the forces of competition and market clearing yield order book dynamics consistent with several stylized facts. It also enables us to study how, in this context, the joint evolution of pricing, orders and cancelations reflect the structure of valuations in the population of investors as well as the trading technology.

The next section presents our model. Section 2 characterizes the equilibrium. Section 3 presents implications from our analysis. Section 4 discusses the robustness of our results. The last section presents our conclusions. Proofs not given in the text are in the appendix, or in the addendum to this paper (Biais and Weill, 2009), which also presents additional useful computations and results and an extension of our model.

markets. Asset liquidity has also been analyzed within the monetary search models of Kiyotaki and Wright (1989) and Lagos and Wright (2005): see, for instance, Lagos (2005), Lester, Postelwaite, and Wright (2009a,b) and Rocheteau (2009).

1 Model

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 are 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 utility 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.[3]

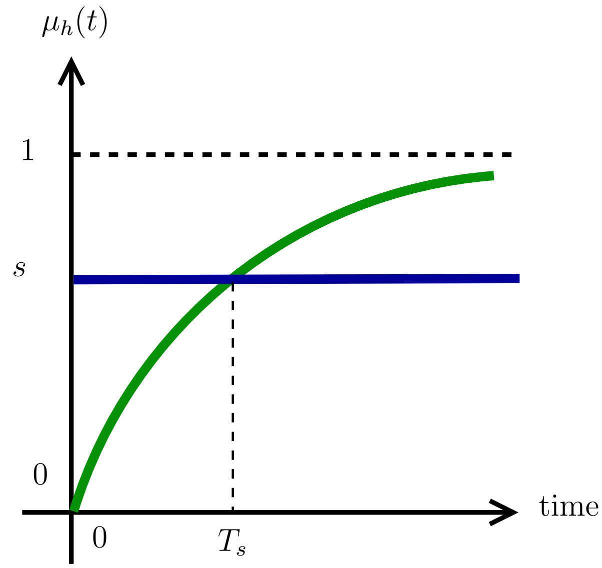

Figure 1: The time path of the fraction $\mu_h(t) = 1 - e^{-\gamma t}$ of high-valuation investors

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:

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

[3] An alternative specification, closer in spirit to Grossman and Miller (1988), would be to assume that investors come from two separate populations: a population of low–valuation investors who initially hold the asset, and a population of high–valuation investors who progressively enter the economy according to the function $\mu_h(t)$. In Section IV of the Addendum (Biais and Weill, 2009), we show that this approach would yield similar results.

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 sums 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. One may wonder why investors do not immediately bid up this predictable price increase? This is because the demand for the asset builds up slowly: on the extensive margin, high-valuation investors cannot hold more than one unit of the asset; and, on the intensive margin, the recovery from the aggregate liquidity shock occurs progressively as investors switch back to high utility flows. Such slow demand build up has been observed in many markets (see e.g., Coval and Stafford, 2007) and reflects "limits to arbitrage."[4]

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), and 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 do not hold the asset are low–utility types. Hence there is a mass $1 - s$ of low utility investors who do not 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 hands of high-valuation investors forever, and the trading volume is zero.

[4]See Lagos and Rocheteau (2009) and Gârleanu (2009) for a comprehensive analysis of position limits in search-and-matching financial markets.

2 Trading with imperfect monitoring

2.1 How the limit order book works

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 investors' utility processes.[5] When an investor contacts the market, she trades through a limit order book. She can place market buy or sell orders immediately executed at the current ask or bid. Otherwise she can place limit sell (resp. buy) orders at prices strictly 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.

A limit order to sell submitted at time t at limit price p is filled after the first time the market price is greater than or equal to p, according to standard price and time priority rules. That is, it is executed at the limit price p *i*), before sell orders at higher prices (price priority) and *ii*) before limit sell orders at the same price submitted after t (time priority). The case of limit buy orders is symmetric. In equilibrium, the number of (market or limit) buy orders executed at time t at the current market price $p(t)$ must equal the number of (market or limit) sell orders filled at that price.

2.2 Equilibrium

We consider competitive equilibria where investors take the price and order book process as given, and respond to this process by placing optimal limit or market, buy or sell orders. In equilibrium, the price and order book process resulting from these orders is that which had been anticipated by the agents, and the market clears at each point in time.

In the remainder of this paper, we restrict our attention to the class of *monotone limit order equilibria* (MLOE), defined as equilibria whose price paths have the following property:

– deterministic, bounded, piecewise continuously differentiable,

[5] In Section 4, we relax this assumption and consider what happens when, with some probability, investors contact the market as soon as their type changes. There we show that the qualitative features of our equilibrium are robust to that extension.

– strictly increasing over some initial time interval $[0, T_f)$, for some $0 \leq T_f \leq \infty$,

– constant at all subsequent times, $[T_f, \infty)$.

These restrictions are satisfied, in particular, by the Walrasian price path decribed above.

2.3 A Unique Candidate Equilibrium

2.3.1 Elementary Properties of an Equilibrium

We first show that a MLOE must have the following elementary properties:

Lemma 1. *In an MLOE, i)* $p(T_f) = 1/r$*, ii)* $T_f < \infty$*, and iii)* $T_f > 0$*.*

To see why points i) and ii) in Lemma (1) hold, note that, if either $T_f < \infty$ and $p(T_f) < 1/r$, or if $T_f = \infty$ and $p(T_f) = \lim_{t\to\infty} p(t) \leq 1/r$, then at each time $p(t) < 1/r$, implying that high-valuation investors find it strictly optimal to buy and hold the asset. Thus, asymptotically all high-valuation investors would hold the asset, which is impossible given that the asset supply s is less than the asymptotic measure of high-valuation investors, $\lim_{t\to\infty} \mu_h(t) = 1$. If, on the other hand $p(T_f) > 1/r$, then for all t large enough no investor would be willing to hold the asset, which also cannot be the basis of an equilibrium.

To see why point iii) holds, note that if $T_f = 0$, the price is immediately equal to $1/r$ at time zero. But, at that time all investors have a low-valuation and thus prefer to sell since the price is $1/r$. Thus, the market supply in a time interval of length Δ around zero is $\rho s \Delta + o(\Delta)$. By the same token, since the measure of high-valuation investors is very close to zero, the demand is of order $o(\Delta)$. So the market cannot clear. Our next lemma states that there is no limit buy order before T_f:

Lemma 2. *In an MLOE, when coming into contact with the market during* $[0, T_f)$*, an investor does not find it optimal to submit a limit buy order.*

This lemma stems from the fact that a limit order to buy at price $p < p(t)$ will never be executed and can therefore be ignored, while a limit buy order at price $p > p(t)$ is executed immediately at price $p(t)$ and is therefore equivalent to a market order.

Given that Lemma 2 allows us to ignore limit-buy orders, we may categorize investors into 6 types. An investor's type includes her utility flow (high "h," or low "ℓ")

and her ownership status (non–owner "n," or owner). We also distinguish asset owners who have not previously placed a limit sell order in the order book ("o") from owners who have previously placed a limit sell order in the book ("b"). Of course, these investors differ in terms of their limit prices, but we need not keep track of these at this stage of the analysis. The set of investor types is, then:

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

2.3.2 The demand of high-valuation investors

First, it is easy to show that, in an MLOE, the asset demand of high-valuation investors has essentially the same form as in the Walrasian case:

Lemma 3. *In an MLOE the following trading plan is strictly optimal for high-valuation investors coming into contact with the market at time $t \in [0, T_f)$:*

> *$\underline{ho}$: hold on to the asset until at least T_f;*
>
> *$\underline{hb}$: cancel any previously placed limit order, then behave as ho;*
>
> *$\underline{hn}$: buy immediately, then behave as ho.*

After T_f, the same plan is weakly optimal.

The trading plan of hn investors reflects that, since the price is increasing towards $1/r$, it is optimal to buy as quickly as possible. Together with investors' ability to submit limit sell orders, the lemma implies:

Lemma 4. *The price path is continuous.*

Indeed if the increasing price process were to jump up at some time $t \in [0, T_f]$, an investor would strictly prefer to submit a limit order to sell immediately after the jump, rather than a (market or limit) sell order just before the jump. Thus, the asset supply just before the jump would be zero. This contradicts market clearing since Lemma 3 shows that the demand is strictly positive at all times $t \in [0, T_f)$.

2.3.3 The supply of low-valuation investors

Next, we study the problem of low-valuation investors. To that end, it is convenient to identify a sell order with its execution time, which we denote by z. Consider an investor who contacts the market at time t. If she places a market sell order, it is immediately executed, i.e., $z = t$. Since the price is strictly increasing over $[0, T_f)$, a limit sell order at price $p(z) \in (p(t), 1/r)$ corresponds to execution time $z > t$. Submitting no sell order or submitting limit prices above $1/r$ results in no execution, which can be identified with execution time $z = \infty$. Lastly, the execution time for the limit price $1/r$ belongs to $[T_f, \infty)$ and is determined by time priority; as will become clear below, in equilibrium this must be equal to T_f. Now let $V_{\ell n}(t)$ be the value (maximum attainable utility) of an ℓn investor and $V_{\ell b}(t, z)$ be that of a ℓb investor with a previously submitted limit-sell order to be executed at time $z > t$. To determine which orders are optimal, we study the function

$$N_\ell(t, z) \equiv V_{\ell b}(t, z) - [V_{\ell n}(t) + p(t)], \tag{3}$$

which is the net utility of submitting a limit sell order to be executed at z, rather than a market sell order. The value obtained with the latter is the current market price $p(t)$, plus continuation value $V_{\ell n}(t)$.

To analyze $N_\ell(t, z)$, consider the derivative $\partial N_\ell(t, z)/\partial z$, which is the marginal value of increasing execution time by dz and behaving optimally thereafter. Note first that the change in execution time is relevant only if the next contact time with the market is greater than z, an event occurring with probability $e^{-\rho(z-t)}$. 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

$$\mathbb{E}_t[\theta(z)]\, dz = (1 - \delta)\, dz + \delta \pi_h(t, z)\, dz, \tag{4}$$

i.e. the investor always enjoys utility flow $1 - \delta$, but on top of this she may enjoy δ if she has switched to 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 $z + dz$ instead of time z. The corresponding net utility is

$$\frac{p(z + dz)}{1 + r dz} - p(z)\, dz = (\dot{p}(z) - rp(z))dz, \tag{5}$$

where $\dot{p}(z)$ is the left derivative of the price path, which is well defined given that the price is continuous (Lemma 4) and piecewise continuously differentiable by assumption. The first term on the right-hand side of (5) is the capital gain of selling at a later time. The second term is the time cost of delaying the sale. Taking the two effects of equations (4) and (5) together, multiplying by the discount factor $e^{-r(z-t)}$ and by the probability $e^{-\rho(z-t)}$ that the execution time is indeed greater than the next contact time, we obtain the marginal value of increasing the execution time z:

$$\frac{\partial N_\ell}{\partial z}(t,z) = e^{-(r+\rho)(z-t)}\left[1-\delta+\delta\pi_h(t,z)+\dot{p}(z)-rp(z)\right]. \tag{6}$$

Integrating over $[t,z]$, and keeping in mind that $N_\ell(t,t)=0$, because a limit order to sell at time $z=t$ clearly yields the same utility as a market sell order at time t, we obtain our next lemma:

Lemma 5. *For a low-valuation investor, the net utility of submitting a limit sell order is*

$$N_\ell(t,z) = \int_t^z \left(1-\delta+\delta\pi_h(t,u)+\dot{p}(u)-rp(u)\right)e^{-(r+\rho)(u-t)}\,du. \tag{7}$$

But when $z \geq T_f$, we have $p(z)=1/r$: thus, equation (6) shows that

$$\frac{\partial N_\ell}{\partial z}(t,z) < 0 \text{ when } z \geq T_f. \tag{8}$$

This has two implications.

First, at any time, low-valuation investors do not find it optimal to submit limit orders with execution time $z > T_f$. This leads to the next lemma, pinning down the execution time of a limit sell order at price $1/r$.

Lemma 6. *In an MLOE, at any time $t < T_f$, a limit sell order at price $1/r$ is executed at time T_f.*

Intuitively, if the execution time were strictly larger than T_f, then low-valuation investors would prefer to submit limit sell orders at a price just below $1/r$, which would be executed just before T_f. But this also implies that the measure of limit sell orders at price $1/r$ would be zero, and therefore the execution time for orders placed at price $1/r$ would be equal to T_f, a contradiction.

The second implication of (8) is that a low-valuation investor holding on to his his asset will find it optimal to submit some limit sell order. Consider, for instance, the submission of a limit order with a limit price just below $1/r$, executed just before T_f, and focus on the case where the investor does not re–contact the market prior to that time, so that his initial order is relevant. Then there are two scenarios: either the investor has switched to high utility by T_f, and since the limit price is $1/r$ he is indifferent between selling and not selling. Or, the investor still has low–utility, in which case he strictly prefers to sell at price $1/r$. This yields our next lemma:

Lemma 7. *A low-valuation investor strictly prefers to submit some (limit or market) sell order executed at some time $z \leq T_f$, rather than simply holding on to her asset until her next contact time with the market ($z = \infty$).*

Based on Lemma 6 and 7, there is no loss of generality in restricting attention to execution times in $[t, T_f]$. Optimal sell orders are, then, characterized by the correspondences:

$$\begin{aligned} \Phi(t) &= \arg\max_{z \in [t, T_f]} N_\ell(t, z) \\ \phi(t) &= \Phi(t) \cap (t, T_f]. \end{aligned}$$

In words, the set $\Phi(t)$ contains all optimal execution times, and the set $\phi(t)$ contains the execution times of optimal, not immediately executed, limit sell orders. Namely, if $t \in \Phi(t)$, then placing a market order is optimal, while if $z \in \phi(t)$, then placing a limit sell at price $p(z)$ is optimal.

This bring us to a key property of our model: investors who contact the market relatively early are more willing to choose relatively late execution times.

Proposition 1. *Consider some contact time t and suppose there is some $z \in \phi(t)$. Then, for all $t' \in (t, z]$, $\Phi(t') \subseteq [t', z]$. In particular, $\Phi(z) = \{z\}$.*

That is, if the investor contacting the market at time t finds it optimal to place a limit order to sell executed at time $z > t$, then the investor contacting the market at a later time $t' > t$ finds it optimal to place an order executed earlier, i.e., between t' and z. In particular, a low-valuation owner contacting the market at time z would place a sell order executed at time z, i.e. a market sell order.

To prove the proposition, we re-scale the net utility $N_\ell(t, z)$ by $e^{-(r+\rho)t}$ and take the derivative with respect to z:

$$\frac{\partial}{\partial z}\left[e^{-(r+\rho)t}N_\ell(t,z)\right] = e^{-(r+\rho)z}\left[1-\delta+\delta\pi_h(t,z)+\dot{p}(z)-rp(z)\right]. \tag{9}$$

This is simply the marginal value of delaying, but measured in time-zero consumption units. Observe that the derivative (9) is *decreasing* in the contact time t; indeed, the only way it depends on the contact time t is through the probability $\pi_h(t,z)$ that during $[t,z]$ there was a switch to high utility. Now, that probability is higher for early t's because early investors have more time to switch to high utility during $[t,z]$. This property makes the low-valuation investor's problem sub-modular: the cross-derivative of the value with respect to the "type" (the contact time) and the "action" (the execution time) is negative. This implies that low types, who contact the market early, choose higher actions, i.e., limit orders with higher limit prices that are executed at later times. Thus, as time passes, sellers place limit orders at lower and lower prices. This is similar to undercutting except that it is not driven by strategic considerations. Rather, it stems from the evolution of the agent's preferences toward liquidity.

2.3.4 Market Clearing

The next proposition describes the implications of market–clearing for the supply stemming from low–valuation agents before T_s:

Proposition 2. *In an MLOE, market clearing implies that the following trading plan must be optimal for low-valuation investors coming into contact with the market at time* $t\in[0,T_s\wedge T_f)$*:*

> $\underline{\ell o}$*: sell with a market order; or place a limit order executed at some time* $z\in\phi(t)$*;*
>
> $\underline{\ell b}$*: cancel previously placed limit order and behave as* ℓo*;*
>
> $\underline{\ell n}$*: stay put.*

In addition, during any measurable set of times $\mathcal{T}\subseteq[0,T_s\wedge T_f)$*, low-valuation investors collectively submit a measure* $\int_{\mathcal{T}}\rho(s-\mu_h(t))\,dt$ *of limit sell orders to be executed at some time* $z\geq T_s\wedge T_f$*.*

It is important to keep in mind that Proposition 2 does not establish the optimality of the trading plan. Instead, it shows what trading plan low-valuation investors must follow in order for the market to clear, given the restrictions imposed by Proposition 1. We will verify optimality below, in Section 2.4.

To prove the proposition, consider any measurable set of times $\mathcal{T} \subseteq [0, T_s \wedge T_f)$. The total measure of high-valuation investors contacting the market during $\mathcal{T}$ is equal to $\int_{\mathcal{T}} \rho\mu_h(t)\, dt$, and the total measure of assets brought to the market is equal to $\int_{\mathcal{T}} \rho s\, dt$. According to Proposition 3 all high-valuation investors must exit the market with one unit of the asset. Since for all $t < T_s$, we have the inequality $\mu_h(t) < s$, this implies that there remains a measure

$$\int_{\mathcal{T}} \rho(s - \mu_h(t))\, dt > 0 \tag{10}$$

of assets that must be held by some other investors. Since limit-buy orders are not submitted, the assets cannot be held by investors who previously posted limit-buy orders. Therefore, the measure of assets given by (10) must be held by the low-valuation investors who contact the market during $\mathcal{T}$. In turn, we know from Lemma 7 that a low-valuation investor who contacts the market and chooses to hold on to an asset finds it optimal to submit some limit sell order. Hence, the measure of assets given by (10) must be held by low valuation agents willing to postpone their sales and correspondingly place limit orders to be executed at later times. Thus, the measure of limit sell orders submitted during $\mathcal{T}$ must be greater than or equal to (10).

In the appendix we show that, at all times in $[0, T_s \wedge T_f)$, limit orders are never executed. Thus, during $\mathcal{T}$ buy orders are executed against market sell orders and low-valuation investors must find it optimal to submit market sell orders. A further implication is that all limit sell orders submitted during $\mathcal{T}$ can be executed only at $z \geq T_s \wedge T_f$. Lastly, this implies that the measure of limit sell orders submitted during $\mathcal{T}$ must be exactly equal to (10). Indeed, the only way there could be a strict inequality is if the measure of assets supplied to the market were greater than $\int_{\mathcal{T}} \rho s\, dt$, which is impossible since limit-sell orders are never executed. The next step is to show:

Lemma 8. *In an MLOE, $T_f > T_s$.*

Suppose instead that $T_f \leq T_s$. By Proposition 2, at all times in $t \in [0, T_f)$, $\phi(t) \neq \emptyset$, i.e., investors submit limit orders. Proposition 2 showed that $\phi(t) \geq T_s \wedge T_f = T_f$, and Lemma 7 showed that $\phi(t) \leq T_f$. Thus, $\phi(t) = T_f$ for all $t \in [0, T_f)$. Now, since any limit order submitted in $[0, T_f)$ is executed with a probability bounded away from zero,[6] This results in a positive measure of orders being executed exactly at time T_f. But this cannot be the basis of an equilibrium since, on the other side of the market, there are no market buy orders and the measure of investors who contact the market exactly at time T_f is equal to zero. Thus we can state our next proposition.

Proposition 3. *In an MLOE the following trading plan is strictly optimal for low-valuation investors coming into contact with the market at (almost all) times $t > T_s$:*

> $\underline{\ell o}$*: sell with a market order;*
>
> $\underline{\ell b}$*: cancel previously placed limit order and behave as ℓo;*
>
> $\underline{\ell n}$*: stay put.*

In addition, during any measurable set of times $\mathcal{T} \subseteq (T_s, T_f)$, a measure $\int_{\mathcal{T}} \rho \left(\mu_h(t) - s\right) \, dt$ of limit sell orders is executed.

Proceeding as above, we know that during $\mathcal{T}$ there is a cumulative measure $\int_{\mathcal{T}} \rho \mu_h(t) \, dt$ of high-valuation investors who all exit the market with one unit of the asset. Since $\mu_h(t) > s$, it follows that the asset demand of high-valuation investors cannot be met using $\int_{\mathcal{T}} \rho s \, dt$, the measure of assets supplied by those investors who contact the market during $\mathcal{T}$. Thus, asset demand must be met by the limit order book, and the measure of limit sell orders executed from the book must be greater than:

$$\int_{\mathcal{T}} \rho \left(\mu_h(t) - s\right) \, dt. \tag{11}$$

Note that this implies in particular that limit orders are executed at almost all times in $t \in (T_s, T_f)$. Thus, by Proposition 1, $\Phi(t) = \{t\}$ at almost all times $t \in (T_s, T_f)$ and market sell orders are strictly optimal for low-valuation investors. One can easily show that this is also true at $t \geq T_f$.

Next, we argue that the measure of executed limit-sell orders must, in fact, be equal to (11). Indeed, there are only two reasons why it could be strictly greater than (11). First, there could be a positive measure of limit buy orders, which is ruled out by

[6] Indeed, a limit sell order submitted at time t corresponding to execution time $z > t$ will be executed if the investor does not manage to re-contact the market and cancel it, which occurs with probability $e^{-\rho(z-t)} \geq e^{-\rho T_f} > 0$.

Lemma 2. Second, low-valuation investors could submit a positive measure of market buy orders. But this is impossible since we just showed that market sell orders are strictly optimal at almost all times in (T_s, T_f).

We proceed with additional properties of the correspondence $\phi(t)$, which specifies the optimal execution times of orders placed at time t:

Proposition 4. *The correspondence $\phi(t)$ is i) non-empty, ii) single valued, ii) strictly decreasing, iv) continuous in $t \in [0, T_s)$ and v) satisfies $\phi(0) = T_f$ and $\lim_{t \to T_s} \phi(T_s) = T_s$.*

The intuition is that, otherwise, there would be either holes or atoms in the limit order book. Holes cannot be the basis of an equilibrium since we know that limit orders must be executed at almost all times in (T_s, T_f). Atoms are not consistent with equilibrium either since there are no limit-buy orders and, at any point in time, the demand originating from those investors who contact the market is a "flow" of measure zero.[7]

The next proposition states the integral equation, which pins down the function $\phi(t)$:

Proposition 5. *For all $t \in [0, T_s)$, $\phi(t)$ is the unique solution of:*

$$\int_t^{\phi(t)} \rho \left(s - \mu_h(u)\right) e^{-\rho(\phi(t) - u)} \, du = 0. \tag{12}$$

In order to prove this Proposition, we let $L(z, \phi(t))$ denote the cumulative number of orders in the book at time z, to be executed before time $\phi(t)$. By construction, $L(z, \phi(t)) = 0$ for $z \leq t$ and for $z \geq \phi(t)$. For $z \in [t, \phi(t)]$, we derive an ordinary differential equation (ODE) for $L(z, \phi(t))$ in several steps.

First, we know from Proposition 2 that the measure of new limit sell orders submitted during $[u, v] \subseteq [0, T_s)$ is equal to $\int_u^v \rho(s - \mu_h(z)) \, dz$. We also know from Proposition 3 that the measure of limit orders executed during any time interval $[u, v] \subseteq (T_s, T_f)$ is $\int_u^v \rho(\mu_h(z) - s) \, dz$. In addition to these new order submissions and executions, all investors with a limit order outstanding who establish a new contact with the market find it strictly optimal to cancel their limit sell order. This results in a cumulative

[7]Note that the continuity of $\phi(t)$ means that limit orders are executed at all times in (T_s, T_f). This implies in turn that the trading plan of Proposition 3 is optimal everywhere instead of almost everywhere.

number of cancelations equal to $\int_u^v \rho L(z,\phi(t))\,dz$. Putting these results together, we find that

$$L(v,\phi(t)) - L(u,\phi(t)) = \underbrace{-\int_u^v \rho L(z,\phi(t))\,dz}_{\text{cancelations}} + \underbrace{\int_u^v \rho\left(s - \mu_h(z)\right)\,dz}_{\text{new submissions and executions}}$$

Taking derivatives yields the ODE:[8]

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

Integrating this ODE with initial condition $L(t,\phi(t)) = 0$, we obtain that $L(z,\phi(t))$ is given by the left-hand side of equation (12). One easily sees that this equation has a unique solution $z > T_s$.

2.3.5 The price path

The first-order necessary condition for the execution time $\phi(t)$ to be optimal at time $t \in [0,T_s)$ is

$$\frac{\partial N_\ell}{\partial z}(t,\phi(t)) = 0 \Leftrightarrow rp(t) = 1 - \delta + \delta\pi_h(\phi^{-1}(t),t) + \dot{p}(t), \tag{13}$$

after applying equation (6) and rearranging. This leads to the following intuition: consider an order that will be filled at time $t \in [T_s, T_f]$. This order was placed previously, at time $\phi^{-1}(t)$. At that time, the investor optimally chooses the price of the limit order, to equalize the cost of waiting a short while longer before execution, $rp(t)$, with the benefits of trading slightly later but at a better price. That benefit is equal to the expected flow of utility from holding the asset, $1-\delta+\delta\pi_h(\phi^{-1}(t),t)$, plus the price increase, $\dot{p}(t)$.

Another optimality condition is that, during $[0,T_s)$, a low-valuation investor must be indifferent between selling immediately with a market sell order, or selling with a limit sell order executed at time $\phi(t)$. Thus, the net utility of submitting a limit-sell order executed at time $\phi(t)$ is $N_\ell(t,\phi(t)) = 0$. Given that $\phi(T_s) = T_s$ and $N_\ell(t,t) = 0$,

[8] Clearly, $L(z,\phi(t))$ is bounded by s so we can bound above the integrand above and obtain that $L(z,\phi(t))$ is Lipchitz with coefficient $2\rho s$, and therefore continuous. Since the integrand is continuous, $L(z,\phi(t))$ is continuously differentiable with respect to z.

this is the same as:

$$\frac{dN_\ell}{dt}(t,\phi(t)) = \frac{\partial N_\ell}{\partial t}(t,\phi(t)) = 0.$$

where the first equality follows from the envelope condition. Taking derivatives in equation (7), using $N_\ell(t,z)=0$, this equation becomes:

$$1-\delta+\dot{p}(t)-rp(t) = \delta\int_t^{\phi(t)} \frac{\partial \pi_h}{\partial t}(t,u)e^{-(r+\rho)(u-t)}\,du. \tag{14}$$

This yields the following intuition. As time elapses during $[0,T_s)$, the "marginal" low-valuation investor must remain indifferent between market sell or limit sell orders. That is, his value of delaying the submission of a market sell order must be equal to his value of delaying the submission of a limit sell order, holding the optimal execution time, $\phi(t)$, constant. The value of delaying submission of a market sell order is given by the left-hand side of equation (14), while the value of delaying submission of a limit sell order is, on the right-hand side, the change in the expected present value of utility flows over the interval $[t,\phi(t)]$. Taken together, this gives:

Proposition 6. *In an MLOE, before T_f, the price path solves the ODEs:*

$$\begin{aligned} t\in[0,T_s], \quad rp(t) &= 1-\delta-\delta\int_t^{\phi(t)} \frac{\partial \pi_h}{\partial t}(t,u)e^{-(r+\rho)(u-t)}\,du+\dot{p}(t) && (15)\\ t\in[T_s,T_f], \quad rp(t) &= 1-\delta+\delta\pi_h(\phi^{-1}(t),t)+\dot{p}(t), && (16)\end{aligned}$$

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

2.3.6 Uniqueness

Taken together, the propositions of this section imply that an MLOE is uniquely characterized: investors' trading plans follow Proposition 3, 2, and 3 with execution times given by Proposition 5. On the other hand, the price is given by Proposition 6. This allows us to state our uniqueness theorem.

Theorem 1. *If an MLOE exists, it is unique.*

2.4 Completing the Existence Proof

The next step is to show that the price path and trading plans are indeed the basis of an equilibrium. To that end, we first note that, by construction of the trading plans, the market clears at each date: supply expressed at the current market price is equal to current demand. Next, we observe that, so long as the price of Proposition 6 is strictly increasing over $[0, T_f]$, we know from Proposition 3 that high-valuation investors' trading plan is optimal at all times, and from Proposition 3 that the low-valuation investors' trading plan is optimal after T_s. Thus, in order to establish that our candidate is indeed an equilibrium, we need only show that the price path of Proposition 6 is strictly increasing over $[0, T_f]$ and that the trading plan of low-valuation investors is optimal over $[0, T_s]$. We prove these properties in the appendix, leading to our existence theorem:

Theorem 2. *There exists an MLOE.*

3 Implications

3.1 Trading volume

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

$$\dot{\mu}_{hn}(t) = -\rho\mu_{hn}(t) + \gamma\mu_{\ell n}(t), \tag{17}$$

where $\mu_{hn}(t)$ and $\mu_{\ell n}(t)$ denote the measure of hn and ℓn investors, and $\dot{\mu}_{hn}(t) = d/dt[\mu_{hn}(t)]$. The first term on the right-hand side of (17) arises because, during a short time interval, there is a flow $\rho\mu_{hn}(t)$ of hn investors who contact the market. All of them submit market buy orders and become ho investors. 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 (17) implies that

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

With the initial condition that $\mu_{hn}(0) = 0$, solving this ODE 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, prior to T_s, instantaneous trading volume is equal to:

$$\mathsf{V}(t) = \rho\mu_{hn}(t) = \gamma(1-s)\frac{\rho}{\rho+\gamma}\left(1-e^{-(\rho+\gamma)t}\right) \tag{18}$$

Equation (18) yields our first implication.

Implication 1. *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 $\mathsf{V}(\cdot)$ in (18) with respect to ρ. This yields

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

which is positive. This yields our second implication:

Implication 2. *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 (18) is symmetric in ρ and γ we also obtain the following implication.

Implication 3. *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 that 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 the flow of new high utility investors.

On the other hand, inspecting (18), one can see that volume goes down with s. This is similar to what happens in the Walrasian case and arises because trading volume reflects the number of low-valuation non-owners who switch to high utility and later contact the market: an increase in s mechanically reduces the number of non-owners, and results in a decrease in volume.

3.2 The number of orders in the book

As in the proof of Proposition 5, we let $L(z, \phi(t))$ denote the stock of limit orders in the book at time z, to be executed before time $\phi(t)$:

$$L(z, \phi(t)) = \int_t^z \rho\,(s - \mu_h(u))\, e^{-\rho(z-u)}\, du.$$

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

$$\mathsf{L}(z) \equiv L(z, \phi(0)) = \int_0^z \rho\,(s - \mu_h(u))\, e^{-\rho(z-u)}\, dz. \tag{19}$$

Substituting the value of $\mu_h(t)$ in equation (19), we get:

$$\mathsf{L}(z) = \int_0^z \rho\left(s - 1 + e^{-\gamma u}\right) e^{-\rho(z-u)}\, du.$$

So the derivative of $\mathsf{L}(z)$ with respect to γ is:

$$\int_0^z -u e^{-\gamma u} e^{-\rho(z-u)}\, du < 0.$$

Thus, we state our next implication.

Implication 4. *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, and the larger the number of orders accumulated in the book.*

The implication suggests that when the liquidity shock is severe so that it takes a long time for investors to recover their willingness to hold the asset, the limit order book is very useful.

3.3 How the limit order market absorbs the liquidity shock

As discussed in Section 3, before time T_s, the flow of low-valuation investors is greater than the flow of high-valuation investors. We interpret this as a buyers' market. In this context, low-valuation investors who own the asset are indifferent between placing limit orders 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.

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

We summarize this in the following implication.

Implication 5. *After the liquidity shock, there are two market regimes: before T_s, there is a buyers' market, in which market orders to sell repeatedly hit the bid quote, while after T_s there is a sellers' 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) and Ellul, Holden, Jain, and Jennings (2007)).

Our results also have implications for the dynamics of the spread and the order book during these two regimes. These are illustrated in Figure 2. The following implication is in line with the figure.

Implication 6. *Just after liquidity shock, the spread is large. Then, limit orders to sell accumulate in the order book, driving down the ask quote, and limit buy orders are placed at higher and higher levels. This results in a decrease in the bid–ask spread. Also, 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 executions of market buy orders lead to a decrease in the stock of limit orders in the book.*

3.4 Technological Change

Over the last 20 years, the technology involved in exchange trading 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 process information, generate alerts on market movements and inform trading and investment

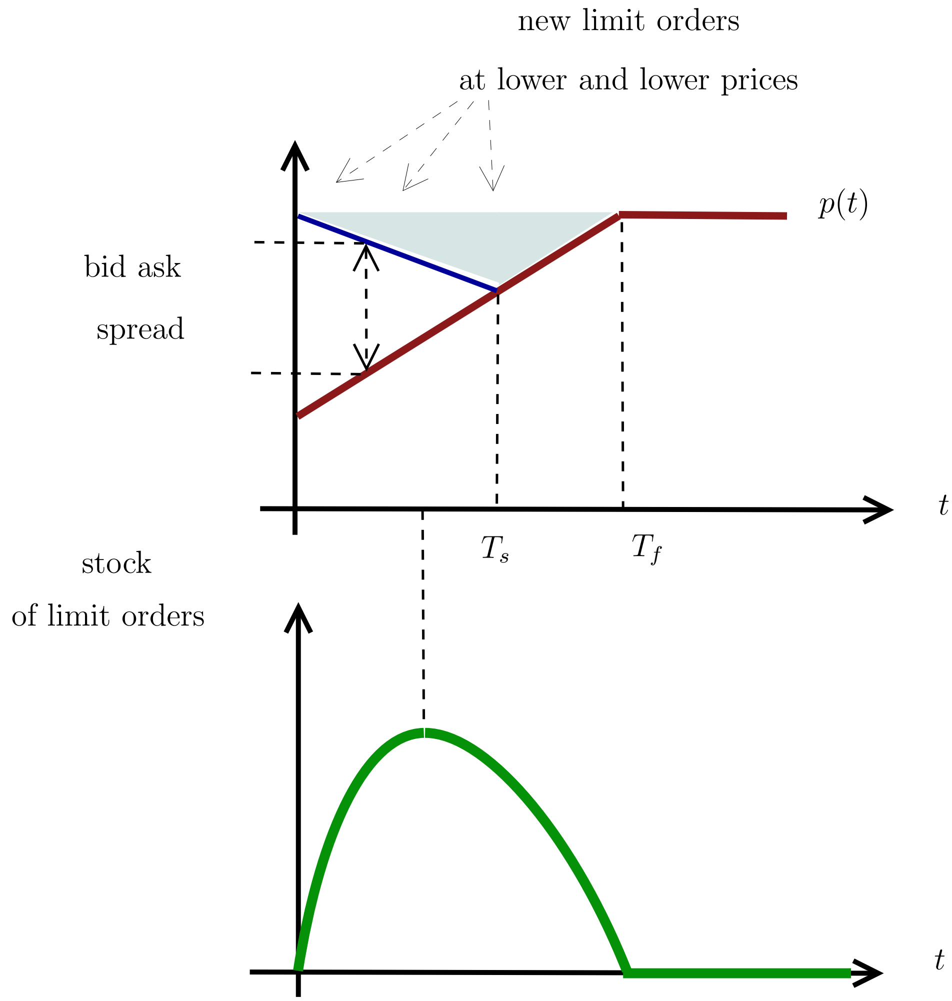

Figure 2: Price and order book dynamics

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 of their study, the NYSE progressively implemented its "autoquote" system, which facilitates the placement of electronic orders, and thus algorithmic trading. These authors find that, as "autoquote" was implemented, the proxy for algorithmic trading (i.e., the ratio of messages to volume) went up.

Our analysis offers a framework to shed light on these technological 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:

Implication 7. *For each $t \in (0, T_s]$, as ρ goes to infinity, the number of orders in the book at time t converges to* $\lim_{\rho\to\infty} \mathsf{L}(t) = s - \mu_h(t)$.

Since, in the limit $\mu_{ho}(t) = \mu_h(t)$, it follows that $s - \mu_h(t)$ is equal to the number of assets in the hands 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 with 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, turning to the behavior of cancelations:

Implication 8. *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 sufficiently large ρ.*

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

$$\mathsf{C}(t) = \rho \mathsf{L}(t).$$

Thus, since the order 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 studied empirically by Hendershott, Jones, and Menkveld (2007):

Implication 9. *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 sufficiently large ρ.*

The flow of messages at time t is

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

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 sell 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:

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

That is, the flow of messages is the sum of the flow $\rho(s-\mu_h(t))$ of new limit orders, the flow $\mathsf{C}(t)$ 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

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

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

4 Robustness

In this section, we examine the robustness of our results to relaxing i) the assumption that preference switching times and market contact times are independent, and ii) the assumption that the intensity of preference switching (γ) is constant.

4.1 Synchronizing contact times with switching times

Investors could benefit from synchronizing switching and contact times. In particular, this could enable them to avoid the execution of a limit sell order after switching to a high valuation. As we show below, our results are robust to allowing for such synchronization.

Consider the following variant of our framework: as in the main model, we let an investor's total intensity of contact with the market be ρ, and a low-valuation investor's switching intensity be γ. We depart from the main model by assuming that a low-valuation investor can partially synchronize his switching time with his contact time. Namely, we assume that, conditional on switching to a high-valuation, a low-valuation investor instantly contacts the market with probability $\varepsilon \in [0, 1]$. To keep low-valuation investors' total contact intensity equal to ρ, we assume that he makes additional contacts with the market, at independent Poisson arrival times with intensity $\rho - \gamma\varepsilon$. Note that if $\varepsilon = 0$, we obtain of course our main model without synchronization. If $\varepsilon = 1$, then there is perfect synchronization and switching times always coincide with contact times. In between, there is partial synchronization.

What is, under partial synchronization, a low-valuation investor's marginal value of increasing his execution time z? As before, the change in execution time is relevant only if the investor does not re-establish contact between the contact time, t, and the execution time, z. Conditional on this event, the marginal value is:

$$\frac{\partial}{\partial z}\left[e^{-(r+\rho)t} N_\ell(t, z)\right] = e^{-(r+\rho)z}\left[1 - \delta + \delta\pi_h^\varepsilon(t, z) + \dot{p}(z) - rp(z)\right],$$

but with $\pi_h^\varepsilon(t, z) \equiv 1 - e^{-\gamma(1-\varepsilon)(z-t)}$. This is the same formula as before except that the probability $\pi_h^\varepsilon(t, z)$ is different. To understand why, recall that $\pi_h^\varepsilon(t, z)$ is the probability that one's utility is high at time z conditional on having a low utility at time t, and *not contacting the market during* $[t, z]$. Thus, the switching intensity that matters for $\pi_h^\varepsilon(t, z)$ is $\gamma(1-\varepsilon) < \gamma$, the arrival intensity of a switching time that is not synchronized with a contact time.

Our key observation is that, although the marginal future value of delaying differs, as long as $\varepsilon < 1$ it remains a decreasing function of the contact time, t. This implies that the result of Proposition 1 goes through identically: low valuation investors find it optimal to undercut each others so long as there is some (possibly arbitrarily small) synchronization problem.

Having established that the order submission problem of low-valuation investors is essentially the same as before, one can solve for an MLOE characterized by two times $T_1 < T_2$ where at any time $t \in [0, T_1]$, low-valuation investors submit limit orders to be executed at some time $\phi(t) \in [T_1, T_2]$, for some strictly decreasing function $\phi(t)$. This is stated in the next proposition, whose proof is sketched in the appendix.

Proposition 7 (Equilibrium with Synchronization)**.** *Suppose $\rho s > \gamma\varepsilon$ and $\varepsilon < 1$. Then there is an MLOE that is identical to that of Theorems 1 and 2 after making the change of variable $\pi_h^\varepsilon(t, z) = 1 - e^{-\gamma(1-\varepsilon)(z-t)}$ and $\mu_h^\varepsilon(t) = 1 - (1 - \gamma\varepsilon/\rho)e^{-\gamma t}$.*

The Proposition shows that our main results continue to hold if investors can synchronize their switching times and their contact times, as long as two conditions are satisfied.

The first condition is that $\rho s > \gamma\varepsilon$. If this condition is not satisfied, the price immediately adjusts to $1/r$ and limit orders are never submitted. The reason is that, with synchronization, the sample of investors in contact with the market is no longer representative of the overall population. Instead, it is biased toward high-valuation investors. The added buying pressure can be large enough to make the number of high-valuation investors in contact with the market larger than the number of assets brought to the market, even at time zero.[9] In that case, high-valuation investors must be made indifferent between holding the asset or not holding it, and the price immediately adjusts to $1/r$. In that case, as we already know from Proposition 4, all low-valuation investors in contact with the market find it optimal to submit market sell orders.

The second condition is that $\varepsilon < 1$. If this condition is not satisfied, then although the equilibrium characterization would also work, low-valuation investors are indifferent regarding the execution time of their limit order. It would then be possible to construct other equilibria where successive limit orders to sell would no longer necessarily undercut each other. Thus, as long as the limit order book at T_1 remains constant, different patterns of evolution of the order book before T_1 are consistent with equilibrium. The

[9]This is, in fact, the meaning of condition $\rho s \leq \gamma\varepsilon$: the flow of assets brought to the market, ρs, is less than the flow of low-valuation investors who switch to highvaluation at time zero, and immediately make contact with the market.

proposition suggests, however, a natural way to select among all these equilibria: by setting ε arbitrarily close to 1, investors face a small synchronization problem and strictly prefer to adopt decreasing order submission strategies.

4.2 Time varying switching rate

Returning to the case where the times at which agents switch to high utility and at which they contact the market are independent, consider now the case where the intensity of switching, γ, is time-varying.

In this generalized framework, let the hazard rate of switching be some positive function $\gamma(t)$, resulting in the aggregate dynamics:

$$\mu_h(t) = 1 - e^{-\int_0^t \gamma(z)\, dz}.$$

Then, one easily shows that the entire uniqueness proof goes through identically; the existence result, however, may fail to hold. This is because, for some specifications of $\gamma(t)$, the price path of Proposition 6 may fail to be increasing. The next corollary provides such a counterexample.

Corollary 1. *Suppose that the hazard rate of switching is $\bar{\gamma}$ initially, during some time interval $[0, \bar{T}]$, and $\underline{\gamma} < \bar{\gamma}$ for $t \geq \bar{T}$. Let $\bar{\gamma} \to \infty$, holding $\bar{\gamma}\bar{T}$ constant. Then, for $\bar{\gamma}$ large enough, the price of Proposition 6 is decreasing for $t \in [0, \bar{T}]$.*

The intuition is that, as time elapses during the initial time interval $[0, \bar{T}]$ without switching, we progressively approach the interval $[\bar{T}, \infty)$, where the switching intensity is much lower. Thus, a low valuation investor becomes more and more pessimistic about her probability of switching in the near future. Now recall that that price must keep these increasingly pessimistic low-valuation investors indifferent between selling immediately and waiting until the best execution time or the next contact time (whichever comes first). Corollary 1 shows that, when $\bar{\gamma}$ is large enough, low-valuation investors' pessimism increases so much that the price stated in Proposition 6 declines.

Note that the effect of Corollary 1 relies on making low-valuation investors more and more pessimistic about their probability of switching before the execution time *or* the next contact time. If $\rho \to \infty$, the next contact time comes much more rapidly than the execution time, and the probability of switching to high utility becomes negligible anyway. Thus, the effect of Corollary 1 does not operate. In that case, we obtain the following corollary.

Corollary 2. *Suppose that $\gamma(t)$ is continuously differentiable and strictly positive. Then, if ρ is large enough, the price of Proposition 6 is strictly increasing.*

Together with our findings discussed above that, even when ρ goes to infinity, the limit order book does not become empty and the bid–ask spread remains strictly positive, Corollary 2 shows that our equilibrium is quite robust to letting ρ become very large.

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 order book builds up. Second, there is a "sellers' market, " in which the flow of buy orders exceeds the flow of buy orders, and trades hit the ask side of the order book. The dynamics generated by our model match the stylized facts on order books, with clustering of activity at the best quotes, undercutting, and serial correlation in order types.

Our model also sheds light on the consequences of the increased 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 future research we plan to extend this framework to endogenize information and market structures.

A Proofs

A.1 Proof of Lemma 1

To keep the appendix short, we omit the proof and relegate it to the Addendum, Section I.1, page 2.

A.2 Proof of Lemma 3

After T_f, since $p(T_f) = 1/r$, a high-valuation investor is indifferent between buying and selling the asset. Next, consider what happens before T_f.

First, consider a high valuation agent who does not own the asset (hn) and contacts the market at time $t \in [0, T_f)$. If he behaves according to the prescribed trading plan of purchasing the asset at t and afterwards follows the optimal policy of keeping the asset, his value is $1/r - p(t)$. To check optimality, by the Bellman principle it suffices to rule out one-stage deviations, whereby an investors deviates once from the prescribed plan, and follows it thereafter. If the high–valuation non–owner deviates once by not submitting a market buy order his value is

$$V_{hn}(t) = \mathbb{E}\left[e^{-r(\tau-t)}\left(\frac{1}{r} - p(\tau)\right)\right], \tag{20}$$

where τ is the next of the agent's contact with the market. This is strictly less than

$$\frac{1}{r} - p(t), \tag{21}$$

because the price is strictly increasing and because of discounting. Note that, by Lemma 2 we do not need to check deviation involving limit buy orders.

Second consider a high-valuation investor who owns the asset. The value of following the prescribed plan of holding on to the asset until at least T_f, is equal to $1/r$. If the agent deviates once and submits an order to sell at time $z \geq t$, his value is:

$$\begin{aligned} & \mathbb{E}_t\left[\int_t^{z\wedge\tau} e^{-r(u-t)}\,du + e^{-r(z\wedge\tau-t)}\left(\left(p(z) + V_{hn}(z)\right)\right)\mathbb{I}_{\{z<\tau\}} + \frac{1}{r}\mathbb{I}_{\{z\geq\tau\}}\right)\right] \\ = \; & \frac{1}{r} + \mathbb{E}_t\left[e^{-r(z\wedge\tau-t)}\left(p(z) + V_{hn}(z)) - \frac{1}{r}\right)\mathbb{I}_{\{z<\tau\}}\right], \end{aligned}$$

where $V_{hn}(z)$ is defined in equation (20). The value above is lower than $1/r$, because, from equations (20)-(21), it follows that $p(z) + V_{hn}(z) - 1/r < 0$ for all $z < T_f$. This shows that a high-valuation investor is strictly better off holding on to her asset until at least T_f. QED

A.3 Proof of Lemma 4

Lemma 9. *The difference between the value derived from placing a limit sell order to be executed at time z and placing a market sell order is*

$$N(t,z) \quad = \quad -p(t) + \mathbb{E}_t\left[\int_t^{z\wedge\tau} \theta(u)e^{-r(u-t)}\,du + e^{-r(z\wedge\tau-t)}p(z\wedge\tau)\right]. \tag{22}$$

The proof goes as follows. Consider a "limit order investor" who submits at time t a limit sell order to be executed at time z, and a "market order investor" who submits at time t a market sell order. These investor have different value (maximum attainable utilities) for three reasons.

1. First, of course, at time t, the market order investor sells immediately so he receives $p(t)$ while the limit order investor receives nothing at that time.
2. Second, the limit order investor continues to hold the asset until either the execution time z or the next contact time τ with the market, enjoying the utility:

$$\int_t^{z\wedge\tau} e^{-r(u-t)}\theta(u)\,du,$$

given any realization of the valuation and contact time processes.

3. Third, there is a difference in continuation utility at $z \wedge \tau$. In order to calculate this difference, consider as before a given realization of the utility and contact time process and distinguish two cases. First, if $z \wedge \tau = z$, i.e. if the execution time comes before the next contact time, then the difference in continuation utility is simply $p(z)$. Indeed, the limit order is executed, the limit-order investor becomes a non-owner with the same valuation and contact time history as the market order investor, so from time z on he has the same continuation utility. If, on the other hand, $z \wedge \tau = \tau$, i.e. if the next contact time comes before the execution time, then the difference in continuation utility is also $p(\tau)$. Indeed at their next contact time the only difference between the limit order and the market order investor is that the limit order investor owns the asset. The limit order investors can always sell his asset and obtain the continuation utility of the market order investor, by adopting the same trading plan. The same is true for the market order investor when he purchases the asset. Thus, the difference in continuation utility at time τ between the two investors is simply the value $p(\tau)$ of the asset.

Thus, given any realization of the contact time and valuation processes, the net utility of a limit order is:

$$\underbrace{-p(t)}_{\text{point 1}} + \underbrace{\int_t^{z\wedge\tau} \theta(u)e^{-r(u-t)}\,du}_{\text{point 2}} + \underbrace{e^{-r(z\wedge\tau-t)}p(z\wedge\tau)}_{\text{point 3}}.$$

Thus, after taking expectation bearing in mind that the distribution of contact time and valuation processes are the same for the limit order and market order investor, we obtain that the net utility of submitting a limit sell order rather than selling is as given in the lemma. QED

Let us now prove that the price path is continuous. First the price is constant for all $t \in (T_f, \infty)$ so, obviously, it is also continuous in that interval. Consider now any time $t \in [0, T_f]$. By construction, in an MLOE, the price is increasing so it cannot have negative jumps; thus, the only thing we need to show is that the price cannot have positive jumps. Towards a contradiction, suppose there were a positive price jump at some time $v \in [0, T_f]$. Consider a small time interval before v where the price is continuous (such an interval must exist given that the price is assumed to be piecewise continuously differentiable). Now take any time u in that interval, and any investor who contacts the market at some $t \leq u$. Let us calculate the net-utility of submitting a sell order at a price just slightly below

$p(v^+)$, which is executed at time v, rather than a sell order executed at time u:

$$\begin{aligned}
N(t,v^+) - N(t,u) &= \mathbb{E}_t\left[\mathbb{I}_{\{u\leq\tau\}}\left(-p(u)e^{-r(u-t)} + \int_u^{v\wedge\tau}\theta(z)e^{-r(z-t)}\,dz + e^{-r(v\wedge\tau-t)}p(v^+\wedge\tau)\right)\right] \\
&\geq \mathbb{E}_t\left[\mathbb{I}_{\{u\leq\tau\}}\left(-p(u)e^{-r(u-t)} + e^{-r(v\wedge\tau-t)}p(v^+\wedge\tau)\right)\right] \\
&= \mathbb{E}_t\left[\mathbb{I}_{\{u\leq\tau\}}e^{-r(u-t)}\mathbb{E}\left[-p(u) + e^{-r(v\wedge\tau-u)}p(v^+\wedge\tau)\,|\,\tau\geq u\right]\right] \\
&= \mathbb{E}_t\left[\mathbb{I}_{\{u\leq\tau\}}e^{-r(u-t)}\right]\times\mathbb{E}\left[-p(u) + e^{-r(v\wedge\tau-t)}p(v^+\wedge\tau)\,|\,\tau\geq u\right] \\
&= e^{-(r+\rho)(u-t)}\left\{-p(u) + \mathbb{E}\left[e^{-r(v\wedge\tau-u)}p(v^+\wedge\tau)\,|\,\tau\geq u\right]\right\} \\
&\geq e^{-(r+\rho)(u-t)}\left\{-p(u) + \Pr(\tau\geq v\,|\,\tau\geq u)e^{-r(v-u)}p(v^+)\right\} \\
&\geq e^{-(r+\rho)v}\left\{-p(v) + e^{-(r+\rho)(v-u)}p(v^+)\right\}.
\end{aligned}$$

The equality between third to the fourth line follows because the only random variable in the expectation is τ, so the conditional expectation $\mathbb{E}\left[-p(u) + e^{-r(v\wedge\tau-u)}p(v^+\wedge\tau)\,|\,\tau\geq u\right]$ is in fact a deterlministic function of u. The last inequality shows that $N(t,v^+) - N(t,u)$ is bounded below by a continuous function of u, independent of the contact time t, with limit $e^{-(r+\rho)v}\{-p(v)+p(v^+)\} > 0$ as u goes to v. Hence, for all execution times u close enough to v, investors strictly prefer to delay their order until v^+ because they foresee the jump. It follows that, in a small time interval to the left of v, there are no limit-sell order nor market sell order executed. On the other hand, since $v < T_f$, we know from Lemma 3 that the demand originating from high-valuation non-owners is positive. To see why, note that the measure of high-valuation non-owners in the overall population is bounded below by:

$$e^{-\rho t}(1-s)\mu_h(t),$$

which is the measure of high-valuation non-owners who never contacted the market at time t. Indeed, $e^{-\rho t}$ is the measure of investors who, at time t, have never contacted the market. Since they never had the opportunity to trade, the proportion of non-owner in this subpopulation stays the same over time so it must be equal to its time-zero value, $1-s$. This explains the second term of the product. Lastly, each of these investors is equally likely to be have valuation at time t, with a probability $\mu_h(t)$. Thus, in any time interval of length Δ before, v, the cumulative demand originating from high valuation investors is bounded below by:

$$\rho e^{-\rho v}(1-s)\mu_h(v)\Delta + o(\Delta).$$

Since we know from above that the supply is zero, it follows that the market does not clear, a contradiction. QED

A.4 Proof of Lemma 5

We start from equation 28 in Lemma 9.

$$
\begin{aligned}
N_\ell(t,z) &= -p(t) + \mathbb{E}_t\left[\int_t^{z\wedge\tau} \theta(u)e^{-r(u-t)}\,du + e^{-r(z\wedge\tau-t)}p(z\wedge\tau)\right] \\
&= \mathbb{E}_t\left[\int_t^{z\wedge\tau} \mathbb{E}_t\left[\theta(u)\right]e^{-r(u-t)}\,du - p(t) + e^{-r(z\wedge\tau-t)}p(z\wedge\tau)\right] \qquad (23)\\
&= \mathbb{E}_t\left[\int_t^{z\wedge\tau} \left(1-\delta+\delta\pi_h(t,u)\right)e^{-r(u-t)}\,du - p(t) + e^{-r(z\wedge\tau-t)}p(z\wedge\tau)\right] \qquad (24)\\
&= \mathbb{E}_t\left[\int_t^{z\wedge\tau} \left(1-\delta+\delta\pi_h(t,u)+\dot p(u)-rp(u)\right)e^{-r(u-t)}\,du\right] \qquad (25)\\
&= \mathbb{E}_t\left[\int_t^{z} \mathbb{I}_{\{u\le\tau\}}\left(1-\delta+\delta\pi_h(t,u)+\dot p(u)-rp(u)\right)e^{-r(u-t)}\,du\right] \\
&= \int_t^{z} \mathbb{E}_t\left[\mathbb{I}_{\{u\le\tau\}}\right]\left(1-\delta+\delta\pi_h(t,u)+\dot p(u)-rp(u)\right)e^{-r(u-t)}\,du \qquad (26)\\
&= \int_0^{z} \left(1-\delta+\delta\pi_h(t,u)+\dot p(u)-rp(u)\right)e^{-(r+\rho)(u-t)}\,du, \qquad (27)
\end{aligned}
$$

where equation (23) follows because of the independence between the contact time and the valuation process, equation (24) follows because of the definition of $\pi_h(t,u)$, equation (25) follows because $p(t)$ is continuous and piecewise continuously differentiable, so we can write that

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

on any interval $[t_1,t_2]$. Equation (26) follows after changing the order of the expectation and the integration sign and noting that, at each time u, $\mathbb{I}_{\{u\le\tau\}}$ is the only random variable in the integrand. Lastly, equation (27) follows from the fact that $\Pr(\tau\ge u\,|\,\tau\ge t) = e^{-\rho(u-t)}$. QED

A.5 Proof of Lemma 6

To keep the appendix short, we omit the proof and relegate it to Addendum, Section I.2, page 3.

A.6 Proof of Proposition 1

Consider any time t and suppose there exists $z \in \phi(t)$. Then, we have that for all $z' > z$,

$$
\begin{aligned}
&0 \le N_\ell(t,z) - N_\ell(t,z') \Leftrightarrow 0 \ge \int_z^{z'} \left(1-\delta+\delta\pi_h(t,u)+\dot p(u)-rp(u)\right)e^{-(r+\rho)(u-t)}\,du \\
\Rightarrow\quad &0 > \int_z^{z'} \left(1-\delta+\delta\pi_h(t',u)+\dot p(u)-rp(u)\right)e^{-(r+\rho)(u-t')}\,du \Leftrightarrow 0 < N_\ell(t',z) - N_\ell(t',z'),
\end{aligned}
$$

for all $t' \in (t,z]$, and where the move from a weak to a strict inequality follows because $\pi_h(t,u)$ is strictly decreasing in t. This means that, for all $t' \in (t,z]$ the execution time z strictly dominates all subsequent execution times, implying that $\Phi(t') \subseteq (t,z]$. Since execution times must be greater than submission times, i.e. $\Phi(t') \subseteq [t',\infty)$, the result follows. In the special case when $t' = z$, then we obtain that $\Phi(z) = \{z\}$, since it is included in $\{z\}$ and cannot be empty (being the argmax of a continuous function over a compact set).

A.7 Proof of Proposition 2

The only two things left to show to establish the proposition are that limit-sell orders are never executed during $(0, T_s \wedge T_f)$ and that low valuation investors find it optimal to submit market and limit sell orders at all times during $[0, T_s \wedge T_f)$.

A.7.1 Limits sell orders are never executed during $(0, T_s \wedge T_f)$

Suppose that, during $(0, T_s \wedge T_f)$, limit orders are executed at all times in some set $\mathcal{T}$. Our goal is to show that $\mathcal{T} = \emptyset$.

By, Proposition 1, for all $z \in \mathcal{T}$, $\phi(z) = \emptyset$. That is, when contacting the market at all times in $\mathcal{T}$, low-valuation investors do not submit limit orders. But we know that limit-sell orders are submitted at almost all times in $[0, T_s \wedge T_f)$. Therefore, the set $\mathcal{T}$ must be of measure zero. Now consider some execution time $z \in \mathcal{T}$, corresponding to a limit order at price $p(z)$, submitted at some time $t < z$. By Proposition 1, at all times $t' \in (t, z)$, investors submit orders with execution times in $[t', z]$. The probability that such a limit order is executed is greater than $e^{-\rho(z-t')}$, the probability that an investor does not get a chance to re-contact the market and cancel his order by time z. Clearly, $e^{-\rho(z-t')} \geq e^{-\rho(z-t)}$. Now, applying the the law of large numbers, it follows that the fraction of limit orders submitted during $[t, z]$ that are executed during $[t, z]$ is greater than $e^{-\rho(z-t)}$. Thus, using equation (10) the total number of limit sell orders submitted *and* executed during $[t, z]$ is greater than:

$$\left[\int_t^z \rho(s - \mu_h(u))\, du\right] e^{-r(z-t)} > 0. \tag{28}$$

Now, since the set of execution times in $(0, T_s \wedge T_f)$ is of measure zero, it follows that the positive measure (28) of limit sell orders is executed during a measure zero set of times, $\mathcal{T}'$. But Lemma 2 showed that investors do not submit limit buy orders, so the cumulative asset demand during $\mathcal{T}'$ is less than the cumulative number of investors who contact the market during $\mathcal{T}'$, which is equal to zero because $\mathcal{T}'$ is of measure zero. Thus we have a positive measure of asset supplied, but at the same time a measure zero of asset demanded, which contradicts market clearing. QED

A.7.2 Low-valuation investors find it optimal to submit market and limit sell orders at all times $t \in [0, T_s \wedge T_f)$

We know that these properties must be true at almost all times in $[0, T_s \wedge T_f)$. Therefore, given any time $t \in [0, T_s \wedge T_f)$, there is a sequence t_k converging towards t such that $\{t_k, z_k\} \subseteq \Phi(t_k)$ for some $z_k > t_k$. Moreover, because limit sell orders are never executed during $(0, T_s \wedge T_f)$, and always executed before T_f (Lemma 7), we know that $z_k \in [T_s \wedge T_f, T_f]$. Since the sequence z_k is bounded, it has a subsequence converging to some (t, z), where $z \geq T_f$. By the Theorem of the Maximum (see, e.g., Theorem 3.6 in Stokey and Lucas, 1989) the correspondence $\Phi(\cdot)$ is upper hemi continuous, and so $(t, z) \in \Phi(t)$. This establishes that, at time t, submitting a market sell order and submitting a limit sell orders are both optimal. QED

A.8 Proof of Proposition 4

Point i) is a re-statement of a finding in Proposition 2: low-valuation investors submit limit sell orders at all times in $[0, T_s)$.

As for point ii), suppose to the contrary that for some $t \in [0, T_s)$, $\phi(t)$ has more than one element, say $z_1 < z_2 \in (T_s, T_f)^2$. Then by Proposition 1, for all $t' \in [0, t)$, $\phi(t') \geq z_2$ and for all $t' \in (t, T_s)$, $\phi(t') \leq z_1$. Thus, there is a "hole" in the limit-order book, in that no limit order are executed during the interval (z_1, z_2). But this contradicts Proposition 3 according to which limit orders have to be executed at almost all times during (z_1, z_2).

For the point iii), we already know from Proposition 1 that $\phi(t)$ is decreasing so the only thing left to show is that it is strictly decreasing. Suppose to the contrary that $\phi(t)$ has a "flat spot," i.e. that it is constant equal to z over some interval $[t, t']$. By Proposition 2 there is a strictly positive measure of limit orders submitted during $[t, t']$, and all these limit orders are executed with a probability bounded away from zero. Thus, there is a positive measure limit orders executed exactly at time z. But this can't be the basis of an equilibrium since, on the other side of the market, there are no market buy order (Lemma 2) and the asset demand originating from investors contacting the market exactly at time z is zero.

For point iv), we note that since $\phi(t)$ is a decreasing function, it has countably many discontinuity points. Now consider any discontinuity point, t. We must have $\phi(t^+) < \phi(t)$ since $\phi(t)$ is decreasing. By Proposition 1, low-valuation investors contacting the market at $t' < t$ submit limit orders executed at times $z' \geq \phi(t)$, and low-valuation investors contacting the market at $t' > t$ submit limit orders executed at times $z' \leq \phi(t^+)$. Thus, the measure of limit orders executed during the time interval $(\phi(t^+), \phi(t))$ is bounded above by the measure of low-valuation investors contacting the market exactly at time t, which is equal to zero. But this is a contradiction since, by Proposition 2, there must be a strictly positive measure of limit orders executed during this time interval.

Finally for point v), we know that $\partial N_\ell / \partial z(t, z) < 0$ for $z > T_f$, so $\phi(t) \leq T_f$ and in particular $\phi(0) \leq T_f$. In addition we cannot have $\phi(0) < T_f$ because otherwise the measure of limit orders executed during $[\phi(0), T_f]$ would be zero, in contradiction with Proposition 3. Similarly, by Proposition 2 we must have $\lim_{t \to T_s} \phi(t) = T_s$ because otherwise no limit order would be submitted at all times close enough to T_s. QED

A.9 Proof of Proposition 6

Because low-valuation investors who contact the market at time $t \in [0, T_s]$ must find it optimal to submit a limit-sell order with execution time $\phi(t)$, it follows that

$$\begin{aligned} \lim_{z \to \phi(t)^-} \frac{N_\ell(t, \phi(t)) - N_\ell(t, z)}{\phi(t) - z} &\geq 0 \\ \lim_{z \to \phi(t)^+} \frac{N_\ell(t, \phi(t)) - N_\ell(t, z)}{\phi(t) - z} &\leq 0 \end{aligned}$$

which gives, after taking left and right derivative in equation (7):

$$1 - \delta + \delta \pi_h(t, \phi(t)) + \dot{p}(\phi(t))^- - rp(\phi(t)) \leq 0 \tag{29}$$

$$1 - \delta + \delta \pi_h(t, \phi(t)) + \dot{p}(\phi(t))^+ - rp(\phi(t)) \geq 0 \tag{30}$$

Note that $p(z)$ is continuous and is assumed to be piecewise continuously differentiable. Thus, the above inequalities have to hold with an equality except may be at countably many kink points. At a kink point, either one of the two above inequality must be strict. For instance, suppose that

$$1 - \delta + \delta\pi_h(t, \phi(t)) + \dot{p}(\phi(t))^- - rp(\phi(t)) > 0.$$

Then, because the price path is piecewise continuously differentiable and $\phi(t)$ is continuous, this strict inequality must also hold for all t' in a neighborhood of t. But, because of the optimality conditions this means that $p(z)$ has an uncountable number of kink points, which is a contradiction. The ODE follows by using the optimality condition (29)-(30) with an equality, and letting $z = \phi(t)$.

Next we turn to the ODE (15) during $[0, T_s]$. At any time $t < T_s$, some low-valuation investors submit limit sell orders to be executed at $\phi(t)$. At the same time, there is a positive demand originating from high-valuation non-owners, so other low-valuation investors must submit market sell orders. Therefore, in an equilibrium a low-valuation investor contacting the market during $[0, T_s)$ must be indifferent between the two, i.e.:

$$N_\ell(t, \phi(t)) = 0 \quad \Leftrightarrow \quad p(t) = \int_t^{\phi(t)} \left[1 - \delta + \pi_h(t, u) + \dot{p}(u) - rp(u)\right] e^{-(r+\rho)(u-t)}\, du. \tag{31}$$

Recall our maintained assumption that $p(t)$ is piecewise continuously differentiable. In any interval where $p(t)$ is continuously differentiable, we can take derivatives with respect to t and obtain the ODE (15). Starting from the terminal condition given by the solution of ODE (16) evaluated at T_f, we can solve the ODE (15) backward, in each interval where $p(t)$ is continuously differentiable. Clearly this yields to a price path that is, in fact, continuously differentiable everywhere. QED

A.10 Proof of Theorem 2

A.10.1 Low-valuation optimality

The only thing left to show is that the conjectured trading strategies are optimal at all time $t \in [0, T_s)$. By construction, investors are indifferent between submitting a market sell order, and a limit sell order executed at time $z = \phi(t)$. The only thing we need to verify is that these two trading strategies achieve the maximum of $N_\ell(t, z)$ over $[t, T_f]$.

Plugging the ODE 16 for the price in $[T_s, T_f]$ into the derivative $\partial N_\ell(t, z)/\partial z$, we find that:

$$\begin{aligned} e^{(r+\rho)(z-t)} \frac{\partial N_\ell}{\partial z}(t, z) &= \delta\Big(\pi_h(t, z) - \pi_h(\phi^{-1}(z), z)\Big) \\ &= \delta\left(\frac{\mu_h(z) - \mu_h(t)}{1 - \mu_h(t)} - \frac{\mu_h(z) - \mu_h(\phi^{-1}(z))}{1 - \mu_h(\phi^{-1}(z))}\right) \\ &= \delta \frac{1 - \mu_h(z)}{(1 - \mu_h(t))(1 - \mu_h(\phi^{-1}(z)))}\Big(\mu_h(\phi^{-1}(z)) - \mu_h(t)\Big). \end{aligned} \tag{32}$$

Since $\phi^{-1}(z)$ is decreasing and $\mu_h(t)$ increasing, it follows that the above expression is strictly positive for $z \in [T_s, \phi(t))$, and strictly negative for $z \in (\phi(t), T_f]$. Therefore, $\phi(t)$ achieves the strict maximum of $N_\ell(t, z)$ over $[T_s, T_f]$: in words, the best execution time in $[T_s, T_f]$ is $\phi(t)$. Turning to $z \in [t, T_s]$, we plug the ODE (15) in the formula (7) for $N_\ell(t,z)$:

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

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

$$\begin{aligned} & \delta \int_t^z e^{-(r+\rho)(v-t)} \pi_h(t,v)\, dv = \delta \int_t^z e^{-(r+\rho)(v-t)} \left(- \int_t^v \frac{\partial \pi_h}{\partial u}(u,v)\, du \right) dv \\ = \quad & -\delta \int_t^z \int_u^z e^{-(r+\rho)(v-t)} \frac{\partial \pi_h}{\partial u}(u,v)\, dv\, du, \end{aligned} \tag{33}$$

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

$$\begin{aligned} & \delta \int_t^z e^{-(r+\rho)(u-t)} \int_u^{\phi(u)} e^{-(r+\rho)(v-u)} \frac{\partial \pi_h}{\partial u}(u,v)\, dv\, du \\ = \quad & \delta \int_t^z \int_u^{\phi(u)} e^{-(r+\rho)(v-t)} \frac{\partial \pi_h}{\partial u}(u,v)\, dv\, du \end{aligned} \tag{34}$$

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

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

which is negative because $\pi_h(u,v)$ is strictly decreasing in its first argument. Thus we have shown that, for all $z \in (t, T_s]$, $N_\ell(t,z) < 0$. This establishes that a low-valuation investor strictly prefers to submit a market sell order than a limit-sell order to be executed at $z \in (t, T_s]$. QED

A.10.2 The price is increasing

We first verify that the price is increasing in $[T_s, T_f]$. We first note that, at $t = T_f^-$, $p(T_f) = 1/r$ and using the ODE (16):

$$\dot{p}(T_s) = 1 - (1-\delta) - \delta \pi_h(0, T_f) = \delta(1 - \pi_h(0, T_f)) > 0.$$

Now, for $t \in [T_s, T_f]$, we note that $\phi^{-1}(t)$ is continuously differentiable (see, in the Addendum, Section II.1 page II.1) so that can differentiate (16). We obtain:

$$\begin{aligned} & r\dot{p}(t) = \frac{d}{dt} \pi_h(\phi^{-1}(t), t) + \ddot{p}(t) \\ \Rightarrow \quad & \dot{p}(t) = \int_t^{T_f} \left[\frac{d}{du} \pi_h(\phi^{-1}(u), u) \right] e^{-r(u-t)}\, du + e^{-r(T_f - t)} \dot{p}(T_f^-) > 0. \end{aligned}$$

Indeed, because $\pi_h(t,u)$ is decreasing in t and increasing in u, and because $\phi^{-1}(u)$ is decreasing, it follows that $d/du\big[\pi_h(\phi^{-1}(u), u)\big] > 0$.

We now turn to the interval $[0, T_s]$ and show that under our assumption that the hazard rate of switching is constant the price is increasing. Using the functional form $\pi_h(t,u) = 1 - e^{-\gamma(u-t)}$ we obtain that $d(t) = \dot{p}(t)$ solves the ODEs:

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

After integrating these two ODEs we obtain that, for $t \in [0, T_s]$:

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

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. QED

A.11 Proof of Implication 7

Lemma 10 (Preliminary result)**.** *Let $f(u)$ be some bounded measurable function, continuous at t. Then*

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

as ρ goes to infinity.

Because of continuity, given any $\varepsilon > 0$ there is some $0 < \eta < t$ such that $|f(u) - f(t)| < \varepsilon/2$ for all $u \in [\eta, t]$. Thus,

$$\begin{aligned} & \left|\int_0^t f(u)\rho e^{\rho(u-t)}\,du - f(t)\right| = \left|\int_0^t (f(u) - f(t))\rho e^{\rho(u-t)}\,du + e^{-\rho t}f(t)\right| \\ \leq \quad & 2\sup|f(u)|\int_0^\eta \rho e^{\rho(u-t)}\,du + \int_\eta^t \rho e^{\rho(u-t)}|f(u) - f(t)|\,du + e^{-\rho t}f(t) \\ \leq \quad & 2\sup|f(u)|\left(e^{\rho(\eta-t)} - e^{-\rho t}\right) + \frac{\varepsilon}{2}\left(1 - e^{\rho(\eta-t)}\right) + e^{-\rho t}f(0) \leq \varepsilon, \end{aligned}$$

for ρ large enough. QED

The result follows from applying the lemma . QED

A.12 Proof of Implication 8

To prove the result we first establish the following lemma.

Lemma 11. *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.

To prove this lemma 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

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

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

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

The result then follows by noting that the first two terms go to zero as ρ goes to infinity, and by applying Lemma 10 to the third term. QED

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

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

Also note that

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

Thus, an application of Lemma 11 implies that $\rho^2 \partial \mathsf{L}/\partial\rho(t)$ goes to $-\dot{\mu}_h(t)$ as ρ goes to infinity. Hence $\rho \partial \mathsf{L}/\partial \rho(t)$ goes to 0 and, $\partial C/\partial\rho(t)$ goes to $s - \mu_h(t)$ which is positive for $t < T_s$. QED

A.13 Proof of Implication 9

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

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

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

$$\begin{aligned}\frac{s-\mu_h(t)+\mathsf{L}(t)+\rho\frac{\partial\mathsf{L}}{\partial\rho}}{\rho\Big(s-\mu_h(t)+\mathsf{L}(t)\Big)} &= \frac{1}{\rho}\frac{s-\mu_h(t)+\mathsf{L}(t)+o(1)}{s-\mu_h(t)+\mathsf{L}(t)} \\ &= \frac{1}{\rho}+o\left(\frac{1}{\rho}\right). \end{aligned} \tag{35}$$

Now using equation (18), we obtain that

$$\begin{aligned}\frac{1}{\mathsf{V}(t)}\frac{\partial\mathsf{V}}{\partial\rho} &= \frac{1}{\rho}-\frac{1}{\rho+\gamma}+\frac{te^{-(\rho+\gamma)t}}{1-e^{-(\rho+\gamma)t}} \\ &= \frac{\gamma}{\rho(\rho+\gamma)}+\frac{te^{-(\rho+\gamma)t}}{1-e^{-(\rho+\gamma)t}} \\ &= o\left(\frac{1}{\rho}\right). \end{aligned} \tag{36}$$

Comparing (35) and (36), it the follows that $\mathsf{M}(t)/\mathsf{V}(t)$ increases for ρ large enough. QED

A.14 Proof of Proposition 7

To construct a MLOE, we use a guess and verify method. We guess is that the equilibrium has the same form as in the main model and is characterized by two times $T_1 < T_2$. The price is strictly increasing over $[0, T_2]$ and constant equal to $1/r$ over $[T_2, \infty)$. At all times t during some initial interval $[0, T_1]$, low-valuation investors submit market sell orders and limit sell orders to be executed at $\phi(t) \in [T_1, T_2]$. At all times during $[T_1, \infty)$, low-valuation investors submit market sell orders. The times T_1, T_2 and the function $\phi(t)$ have to be determined in equilibrium.

We then know from the main model (Proposition 3) that high-valuation investors submit market buy orders when contacting the market during $[0, T_2)$, and are indifferent between buying or not afterwards. As in the main model, the optimal order of a low-valuation investor is determined by the net-utility $N_\ell(t, z)$ of submitting at time t a limit-sell order executed at time z, rather than a market-sell order. The derivation is the same as in the main model except for the fact that switching and contact times are no longer independent. Formally, we let $\tau_{\rho-\gamma\varepsilon}$ be the next contact time without simultaneous switching, and $\tau_{\gamma\varepsilon}$ be the next contact time with simultaneous switching, and $\tau = \tau_{\rho-\gamma\varepsilon} \wedge \tau_{\gamma\varepsilon}$ be the next contact time. Proceeding as in the Proof of Lemma 5, we find that:

$$\begin{aligned}
N_\ell(t,z) &= \mathbb{E}_t\left[\int_t^z \mathbb{I}_{\{u\le\tau_{\rho-\gamma\varepsilon}\}}\mathbb{I}_{\{u\le\tau_{\gamma\varepsilon}\}}\left(\theta(u)+\dot p(u)-rp(u)\right)e^{-r(u-t)}\,du\right] \\
&= \int_t^z \mathbb{E}_t\left[\mathbb{I}_{\{u\le\tau_{\rho-\gamma\varepsilon}\}}\right]\text{Proba}_t\left[u\le\tau_{\gamma\varepsilon}\right]\left(\mathbb{E}_t\left[\theta(u)\,|\,u\le\tau_{\gamma\varepsilon}\right]+\dot p(u)-rp(u)\right)e^{-r(u-t)}\,du \\
&= \int_t^z e^{-(r+\rho)(u-t)}\Big(1-\delta+\delta\left[1-e^{-\gamma(1-\varepsilon)(u-t)}\right]+\dot p(u)-rp(u)\Big)\Big)\,du \\
&= \int_t^z e^{-(r+\rho)(u-t)}\left(1-\delta+\delta\pi_h^\varepsilon(t,z)+\dot p(u)-rp(u)\right)\right)\,du
\end{aligned}$$

where the second line follows from the independence between $\tau_{\gamma\varepsilon}$ and $\tau_{\rho-\gamma\varepsilon}$. Note that the expression is the same as in the main model after the change of variable of the proposition. Given $\pi_h^\varepsilon(t,z)$ and given the strictly decreasing function $\phi(t)$ mapping $[0,T_1]$ onto $[T_1,T_2]$, we construct the price path using the ODE of Proposition 6. Going through the same argument as in the main model, one shows easily that, given this price, the trading plans of investors with low-valuation are optimal. Lastly, given the functional form of $\pi_h^\varepsilon(t,z)$, we know from Section A.10.2 that the price path is indeed strictly increasing over $[0,T_2]$.

All what's left to determine, then, is the function $\phi(t)$ and the two times T_1 and T_2. To that end, we note that at any time $t\in[0,T_2]$, there is a flow

$$\rho\mu_{hn}(t)+\gamma\varepsilon\mu_{\ell n}(t), \tag{37}$$

of high-valuation investors who contact the market and demand one unit of the asset. In particular the second term is the flow of low-valuation investors who switch to high and immediately contact the market. On the other side of the market, we have a flow

$$(\rho-\gamma\varepsilon)\left(\mu_{\ell o}(t)+\mu_{\ell b}(t)\right), \tag{38}$$

of low-valuation investors who contact the market with one unit of the asset. It is important to keep in mind that the relevant intensity of contact for equation (38) *is not* ρ but $\rho-\gamma\varepsilon<\rho$. This is because of synchronization: some low-valuation owners contact the market but instantly switch to high, so they do not submit any sell order. Subtracting (38) from (37), we find that the net buy order flow is

$$\begin{aligned}
&(\rho-\gamma\varepsilon)\left(\mu_{hn}-\mu_{\ell o}(t)-\mu_{\ell b}(t)\right)-\gamma\varepsilon\left(\mu_{\ell n}(t)+\mu_{hn}(t)\right) \\
=\;&(\rho-\gamma\varepsilon)\left(\mu_h(t)-s\right)+\gamma\varepsilon\left(1-s\right) \\
=\;&\rho\mu_h(t)+\gamma\varepsilon\left(1-\mu_h(t)\right)-\rho s.
\end{aligned}$$

The first term is the flow of high-valuation investors who contact the market at each time. The second term is the flow of low-valuation investors who switch to high and immediately contact the market. The third term is the flow of assets brought to the market by all investors. The only difference with the main model is, therefore, the second term. It arises because of synchronization: the market is no longer representative of the overall population, but is instead biased towards high-valuation investors. This calculation allows us to write the net buy order flow as:

$$\rho\left(\mu_h^{\varepsilon}(t) - s\right) \text{ where } \mu_h^{\varepsilon}(t) = 1 - \left(1 - \frac{\gamma\varepsilon}{\rho}\right) e^{-\gamma t}.$$

Given that limit orders are canceled with intensity ρ, the construction of $\phi(t)$, T_1 and T_2 is exactly as in the main model. That is, T_1 solves $\mu_h^{\varepsilon}(T_1) = s$, and $\phi(t)$:

$$\int_t^{\phi(t)} \rho(s - \mu_h(z)) e^{\rho(z-\phi(t))}\, dz = 0,$$

for all $t \in [0, T_1]$. Finally, observing that $T_2 = \phi(0)$ completes the proof. QED

A.15 Proof of Corollary 1

Let $\tilde{\mu}_h(u)$ be the solution of $\dot{\tilde{\mu}}_h(u) = \underline{\gamma}\left(1 - \tilde{\mu}_h(u)\right)$ with initial condition $1 - e^{-\bar{\gamma}\bar{T}}$. Then the measure of high valuation investors at time t is $1 - e^{-\bar{\gamma}t}$ if $t \leq \bar{T}$ and $\tilde{\mu}_h(t - \bar{T})$ if $t \geq \bar{T}$. In particular, if we let $\tilde{T}_s$ solve $\tilde{\mu}_h(\tilde{T}_s) = s$, we have that $T_s = \bar{T} + \tilde{T}_s$. For all $t \leq \bar{T}$, the probability $\pi_h(t, u)$ of switching to high in the interval $[t, u]$, is given by:

$$\begin{aligned} u \leq \bar{T} \quad &: \quad \pi_h(t, u) = 1 - e^{-\bar{\gamma}(u-t)} \\ u \geq \bar{T} \quad &: \quad \pi_h(t, u) = 1 - e^{-\bar{\gamma}(\bar{T}-t) - \underline{\gamma}(u-\bar{T})}. \end{aligned}$$

For $t \geq \bar{T}$ we have, as before, $\pi_h(t, u) = 1 - e^{-\underline{\gamma}(u-t)}$. Now consider any $t \leq \bar{T}$. From equation (15), we have that:

$$\begin{aligned} \dot{p}(t) &= rp(t) - (1 - \delta) + \delta \int_t^{\phi(t)} \frac{\partial \pi_h}{\partial t}(t, u) e^{-(r+\rho)(u-t)}\, du \\ &\leq \delta + \delta \int_t^{\tilde{T}_s} \frac{\partial \pi_h}{\partial t}(t, u) e^{-(r+\rho)(u-t)}\, du \end{aligned}$$

where the second line follows because, in a MLOE, $rp(t) \leq 1$, because $\partial \pi_h / \partial t < 0$, and $\tilde{T}_s \leq T_s \leq \phi(t)$. Now using the explicit expression for $\pi_h(t, u)$ above, we find that:

$$\begin{aligned} &\int_t^{\tilde{T}_s} \frac{\partial \pi_h}{\partial t}(t, u) e^{-(r+\rho)(u-t)}\, du \\ = \quad & -\frac{\bar{\gamma}}{r + \rho + \bar{\gamma}} \left[1 - e^{-(r+\rho+\bar{\gamma})(\bar{T}-t)}\right] - \frac{\bar{\gamma} e^{-\bar{\gamma}(\bar{T}-t)}}{r + \rho + \underline{\gamma}} \left[1 - e^{-(r+\rho+\underline{\gamma})(\tilde{T}_s - \bar{T})}\right] \\ \leq \quad & -\frac{\bar{\gamma} e^{-\bar{\gamma}\bar{T}}}{r + \rho + \underline{\gamma}} \left[1 - e^{-(r+\rho+\underline{\gamma})\tilde{T}_s}\right] \to -\infty \end{aligned}$$

as $\bar{\gamma} \to \infty$, holding $\bar{\gamma}\bar{T}$ constant. Plugging this back into the expression for $\dot{p}(t)$ it follows that, for $\bar{\gamma}$ sufficiently large, $\dot{p}(t) < 0$. QED

A.16 Proof of Corollary 2

We now show that the price is increasing in $[0, T_f]$ when ρ is large enough. First of all, the Walrasian price is $p^*(t) = 1/r$ for $t \geq T_s$, and

$$p^*(t) = \frac{1}{r} - \frac{\delta}{r}\left(1 - e^{-r(T_s - t)}\right),$$

for $t < T_s$. The price when $\rho < \infty$ is denoted by $p(t)$. The idea of the proof is to that $p(t)$ and its first derivative, $\dot{p}(t)$, both converges *uniformly* towards the Walrasian price as ρ goes to infinity. Since $\dot{p}^*(t) > 0$ over $[0, T_s)$, this implies that, if ρ is large enough, $\dot{p}(t) > 0$ for all $t \in [0, T_s)$.

A.16.1 Uniform convergence of $p(t)$.

First consider $t \in [T_s, T_f]$.

$$p(t) = \int_t^{T_f} \left[1 - \delta + \delta \pi_h(\phi^{-1}(u), u)\right] e^{-r(u-t)}\, du + \frac{1}{r} e^{-r(T_f - t)}.$$

Thus,

$$\begin{aligned} 0 \leq p^*(t) - p(t) &= \delta \int_t^{T_f} \left[1 - \pi_h(\phi^{-1}(u), u)\right] e^{-r(u-t)}\, du \\ &\leq \delta \int_t^{T_f} e^{-r(u-t)}\, du \leq \frac{\delta}{r}\left[1 - e^{-r(T_f - t)}\right] \leq \frac{\delta}{r}\left[1 - e^{-r(T_f - T_s)}\right]. \end{aligned}$$

Since, for $t \geq T_f$, we have $p(t) = 1/r = p^*(t)$, we obtain that

$$|p(t) - p^*(t)| \leq \frac{\delta}{r}\left[1 - e^{-r(T_f - T_s)}\right],$$

for all $t \geq T_s$. Given that $T_f = \phi(0)$ goes to T_s (see in the Addendum, Section II.2.4, page 8), the convergence is uniform for all $t \geq T_s$. Next, consider some $t \leq T_s$. We have:

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

where

$$\eta(t) \equiv -\delta \int_t^{\phi(t)} \frac{\partial \pi_h}{\partial t}(t, u) e^{-(r+\rho)(u-t)}\, du.$$

Taking the difference with $p^*(t)$, we obtain that

$$|p(t) - p^*(t)| = e^{-r(T_s - t)} |p(T_s) - p^*(T_s)| + \int_t^{T_s} \eta(u) e^{-r(u-t)}\, du$$

Now

$$\frac{\partial \pi_h}{\partial t}(t, u) = \frac{\partial}{\partial t}\left[\frac{\mu_h(u) - \mu_h(t)}{1 - \mu_h(t)}\right] = -\frac{1 - \mu_h(u)}{\left[1 - \mu_h(t)\right]^2} \dot{\mu}_h(t).$$

Since $\mu_h(t)$ is continuously differentiable, $\partial\pi_h/\partial t$ can be bounded above over $[0, \phi(0)]$ by some constant K. Note that, since $\phi(0)$ decrease in ρ (see in the Addendum, Section II.2 page 7) the bound will also work for any larger ρ. We obtain that

$$0 \leq \eta(t) \leq \frac{\delta K}{r+\rho}\left(1 - e^{-(r+\rho)(\phi(t)-t)}\right) \leq \frac{\delta K}{r+\rho}.$$

Plugging this into the expression above we obtain

$$|p(t) - p^*(t)| \quad \leq \quad e^{-r(t-T_s)}|p(T_s) - p^*(T_s)| + \frac{\delta K}{r(r+\rho)}\left(1 - e^{-r(T_s-t)}\right).$$

Since $p(T_s) \to p^*(T_s) = 1/r$, this establishes the uniform convergence of $p(t)$ towards $p^*(t)$ over the interval $[0, T_s]$.

A.16.2 Uniform convergence of $\dot{p}(t)$ over $[0, T_s)$.

To see that $\dot{p}(t)$ converges uniformly towards its limit $rp^*(t) - (1-\delta)$, note that

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

so

$$|\dot{p}(t) - \dot{p}^*(t)| \quad \leq \quad r|p(t) - p^*(t)| + |\eta(t)| \leq r|p(t) - p^*(t)| + \frac{\delta K}{r+\rho},$$

which converges uniformly towards $\dot{p}^*(t)$ as ρ goes to infinity.

A.16.3 The price is increasing for ρ large enough.

Indeed, recall that

$$\dot{p}^*(t) = \delta\left(1 - e^{-r(T_s-t)}\right) > \delta\left(1 - e^{-rT_s}\right) > 0,$$

is uniformly bounded away from zero over $[0, T_s)$. Since $\dot{p}(t)$ converges to $\dot{p}^*(t)$ uniformly for $t \in [0, T_s)$, the result follows. QED

References

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

Bruno Biais and Pierre-Olivier Weill. Addendum to “liquidity shocks and order book dynamics”. Working paper, TSE and UCLA, 2009.

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.

Shane A. Corwin and Jay F. Coughenour. Limited attention and the allocation of effort in securities trading. *Journal of Finance*, 63:3031–3067, 2008.

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

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

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

Darrell Duffie and Gustavo Manso. Information diffusion in large population. *American Economic Review Papers and Proceedings*, 97:203–209, 2007.

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

Darrell Duffie, Nicolae Gârleanu, and Lasse H. Pedersen. Over-the-counter markets. *Econometrica*, 73:1815–1847, 2005.

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

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.

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

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

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.

Nicolae Gârleanu. Portfolio choice and pricing in illiquid markets. *Journal of Economic Theory*, 144:532–564, 2009.

Mikhail Golosov, Guido Lorenzoni, and Aleh Tsyvinski. Decentralized trading with private information. Working Paper, MIT and Yale, 2008.

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.

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.

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

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

Joel Hasbrouck and Gideon Saar. Technology and liquidity provision: the blurring of traditional definitions. *Journal of financial Markets*, 12:143–172, 2009.

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

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

Nobuhiro Kiyotaki and Randall Wright. On money as a medium of exchange. *Journal of Political Economy*, 97:927–954, 1989.

Ricardo Lagos. Asset prices and liquidity in an exchange economy. Working paper, NYU, 2005.

Ricardo Lagos and Guillaume Rocheteau. Liquidity in asset markets with search frictions. *Econometrica*, 77:403–426, 2009.

Ricardo Lagos and Randall Wright. A unified framework for monetary theory and policy analysis. *Journal of Political Economy*, 113:463–484, 2005.

Ricardo Lagos, Guillaume Rocheteau, and Pierre-Olivier Weill. Crashes and recoveries in illiquid markets. Working Paper, NYU, UCI, UCLA, 2007.

Benjamin Lester, Andrew Postelwaite, and Randall Wright. Liquidity and information: I. Working Paper, UWO and UPenn, 2009a.

Benjamin Lester, Andrew Postelwaite, and Randall Wright. Liquidity and information: Ii. Working Paper, UWO and UPenn, 2009b.

Gara Mínguez Afonso. Liquidity and congestion. Working paper, LSE, 2008.

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

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.

Guillaume Rocheteau. A monetary approach to asset liquidity. Working paper, UCI, 2009.

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

Nancy L. Stokey and Robert E. Lucas. *Recursive Methods in Economic Dynamics.* Harvard University Press, Cambridge, 1989.

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

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

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

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

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

Addenda to "Liquidity Shocks and Order Book Dynamics"

Bruno Biais and Pierre-Olivier Weill

May 20, 2009

The following addenda contain supplementary materials.

Addendum I provides some proofs omitted in the paper.

Addendum II provides additional results about the execution time function, $\phi(t)$.

Addendum III provides a step-by-step derivation of the ODEs governing how distribution of types evolves over time.

Addendum IV provides an extension of our model with an additional exogenous inflow of buyers.

I Omitted Proofs

I.1 Proof of Lemma 1

Point *i*) **and** *ii*)**.** First, suppose either that $T_f < \infty$ and $p(T_f) < 1/r$, or that $T_f = \infty$ and $p(T_f) \leq 1/r$. We show that a high valuation agent who does not own the asset (hn) and contacts the market at time $t \in [0, T_f)$ will find it optimal to buy the asset with a market order, and hold it forever after. The value of following this plan is $1/r - p(t)$. To check optimality, by the Bellman principle it suffices to rule out one-stage deviations, whereby a hn investor deviates once from the prescribed plan, and follows it thereafter. If the investors deviates once by not submitting a market buy order his value is

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

where τ is the next of the agent's contact with the market. This is lower than

$$\frac{1}{r} - p(t), \tag{I.2}$$

because the price is strictly increasing and because of discounting. Checking deviation involving limit buy orders (that we have not ruled out at this stage of the analysis), amounts to replace τ by $\min\{\tau, z\}$, where z is the execution time of the order. Clearly, the same argument applies.

Next, suppose that $p(T_f) > 1/r$, and $T_f \leq \infty$. Then for all t large enough, $p(t) > 1/r$. We now show that, for all t large enough, all owners sell their asset at their first contact time with the market, and non-owners don't buy. Indeed the value $V_b(t, z)$ ($V_n(t, z)$) of an owner (non-owner) at time t with a limit order to sell (buy) that is expected to be executed at some time $z > t$, who behave according to this prescribed plan is:

$$\begin{aligned} V_o(t,z) &= \mathbb{E}_t\left[\int_t^{z\wedge\tau} \theta(u)e^{-r(u-t)}\,du + e^{-r(z\wedge\tau-t)}p(T_f)\right] \\ &= p(T_f) + \mathbb{E}_t\left[\int_t^{z\wedge\tau} \left(\theta(u) - rp(T_f)\right)e^{-r(u-t)}\,du\right] \end{aligned} \tag{I.3}$$

$$V_n(t,z) = \mathbb{E}_t\left[e^{-r(z-t)}\mathbb{I}_{\{z<\tau\}}\left(V_o(z,\infty) - p(T_f)\right)\right], \tag{I.4}$$

where the conditioning information is the filtration generated by the investor's valuation and contact time history. Note that, by setting $z = \infty$, we obtain the value of an investor with no limit order. To check the optimality of the owner's plan, by the Bellman principle it suffices to check one-stage deviation, whereby an investors deviates once from the prescribed plan, and follows it thereafter. In the case of an owner, we need to check that the value of selling now, $p(T_f)$, is greater than the value $V_o(t, z)$ of holding on to the asset and submit a sell order with execution time $z > t$, which is true by equation (I.3) since $\theta(u) \leq 1 < rp(T_f)$. In case of a non-owner, we need to check that the value of not having any limit buy order outstanding, $V_n(t, \infty) = 0$, is greater than the value of submitting a limit buy order executed at $z < \infty$. This is also true since equation (I.3) and (I.4) jointly imply that $V_n(t, z) < 0$ for all $z < \infty$. Therefore, if $p(T_f) > 1/r$ starting at time T_f all investors who own sell their asset at their first contact time with the market. Therefore, asymptotically, the measure of non-owner must converge to 0, which is impossible since the measure of asset, s, is strictly less than one. QED

Point *iii*)**.** We first show the following preliminary result: after time T_f, a low-valuation investor who contacts the market finds it strictly optimal to submit market sell orders. Indeed, using equation (I.3) and (I.4) applied to low-valuation investors when $p(t) = p(T_f) = 1/r$, we find

$$V_{\ell o}(t, z) = \frac{1}{r} + \mathbb{E}\left[\int_t^{z \wedge \tau} (\theta(u) - 1)\, e^{-r(u-t)}\, du\right] \tag{I.5}$$

$$V_{\ell n}(t, z) = \mathbb{E}\left[e^{-r(z-t)} \mathbb{I}_{\{z < \tau\}} \left(V_{\ell o}(z, \infty) - \frac{1}{r}\right)\right]. \tag{I.6}$$

Therefore, $V_{\ell o}(t, z) < 1/r = V_{\ell o}(t, 0)$ for all $z > t$ because $\theta(u) = 1 - \delta < 1$ with strictly positive probability over $[t, z \wedge \tau]$. Thus, selling immediately strictly dominates submitting a limit sell order with execution time z. Plugging this inequality back into (I.6) we find that $V_{\ell n}(t, z) < 0$ for all $z < \infty$, meaning that submitting an order to buy executed at any time $z > t$ is not optimal. Having established that all low-valuation investors want to sell, we can apply the argument shown in the text. QED

I.2 Proof of Lemma 6

For each time $t < T_f$, let $z_f(t)$ be the execution time of a limit sell order with limit price $p(T_f) = 1/r$. Keep in mind that order execution follows price and time priority. That is, a limit sell order submitted at time t is executed at the first time such that *i*) the market price is greater than $1/r$, and *ii*) all limit sell orders at price $1/r$ submitted before t have been executed.

Therefore, because of the price priority rule *i*), we must have that $z_f(t) \geq T_f$. We translate the time priority rule *ii*) into the requirement that $z_f(t)$ is increasing, and constant over any time interval when no limit sell orders at price $1/r$ are submitted. The time priority rule also states that a limit sell order submitted at time t has to be executed at the first time after all limit sell orders at price $1/r$ submitted before t are executed. We translate this into the requirement that $z_f(t)$ has to be continuous except at times when investors submit a strictly positive measure of limit sell orders at price $1/r$. Intuitively, we require that an order submitted "just after" time t must be executed "just after" $z_f(t)$, i.e. just after an order submitted at time t, unless there is an atom of limit sell orders submitted at time t. Formally, let $m_f(t)$ be the cumulative measure of limit sell order at price $1/r$ submitted before time t. Then, we require that:

$$\text{if } z_f(t^+) > z_f(t) \text{ then } m_f(t^+) > m_f(t). \tag{I.7}$$

We first prove that in our setup:

Lemma 12. *In a MLOE, the function $m_f(t)$ is continuous. Therefore, requirement (I.7) implies that $z_f(t)$ is continuous.*

Indeed, consider any to times $t_1 < t_2 < T_f$. We must have that:

$$-(1 - e^{-\rho(t_2 - t_2)})m(t_1) \leq m(t_2) - m(t_1) \leq \int_{t_1}^{t_2} \rho \, du. \tag{I.8}$$

The left inequality is the largest possible decrease, during $[t_1, t_2]$, in the number of limit sell orders at price $1/r$. It corresponds to the "worse case scenario" when all investors who previously submitted these orders cancel them at their first contact time with the market, and no additional limit sell orders are submitted at price $1/r$. The right inequality is the largest possible increase in the number of limit sell orders at price $1/r$. It correspond to the "best case scenario" when all investors who contact the market between t_1 and t_2 submit limit sell orders at price $1/r$, and no limit sell orders at price $1/r$ are canceled. Taking the limit $t_2 \to t_1$, we obtain that $m_f(t)$ is continuous.

Now turning to the proof of Lemma 6, suppose that there is some $t < T_f$ such that $z_f(t) > T_f$. Consider the set $\mathcal{T}_t$ of times less than t such that some low-valuation investors submit limit sell orders at price $1/r$. Then, there must be some $v \in \mathcal{T}_t$ such that $z_f(v) > T_f$. Otherwise, by continuity, $z_f(\sup \mathcal{T}_t) = T_f$. Moreover, since no limit sell orders are submitted in the interval $[\sup \mathcal{T}_t, t]$, we must have that $z_f(t) = z_f(\sup \mathcal{T}_t) = T_f$, which is a contradiction. But then we have found some time v such that investors submit limit-sell orders at price $1/r$ even though the execution time is $z_f(v) > T_f$. This contradicts optimality since $\partial N_\ell / \partial z(t, z) < 0$ at $z = z_f(t)$, and that, by submitting a limit sell order at a price smaller but arbitrarily close to $1/r$, an investor can be executed at a time arbitrarily close to T_f. QED

II Additional results about the function $\phi(t)$

In this section we show:

Proposition 8. *The function $\phi(t;\rho,s,\mu_h)$ is: strictly decreasing and continuously differentiable in $t\in[0,T_s]$, strictly decreasing in ρ, strictly increasing in s, is strictly decreasing in $\mu_h(\cdot)$.*

The last Property means that with a "faster" recovery path $\mu_h^1(t)\geq\mu_h^2(t)$, $\phi(t)$ is lower.

II.1 Monotonicity and continuous differentiability

We start by writing:

$$L(t,u)=\int_t^u h(z)\,dz=0,$$

where $h(z)=\rho\,(s-\mu_h(z))\,e^{\rho z}$. Because $\lim_{t\to\infty}\mu_h(t)>s$, it follows that $L(t,u)$ goes to minus infinity as u goes to infinity. Because $\mu_h(t)>s$ for all $t\in[T_s,\infty)$, it follows that $L(t,u)$ is decreasing for all $u\in[T_s,\infty)$. Keeping in mind that $\mu_h(t)<s$ for all $t\in[0,T_s]$, we also have $L(t,T_s)\geq 0$. Thus, for all $t\in[0,T_s]$, there is a unique $\phi(t)\geq T_s$ such that $L(t,\phi(t))=0$.

For all $t<T_s$, $\phi(t)>T_s$ so $\partial L/\partial u=e^{\rho\phi(t)}\,(s-\mu_h(\phi(t)))<0$ at $u=\phi(t)$. Thus, an application of the Implicit Function Theorem (IFT) shows that the function $\phi(t)$ is continuously differentiable over $(0,T_s)$, with a derivative that is equal to:

$$\phi'(t)=\frac{h(t)}{h(\phi(t))}. \tag{II.1}$$

Note that because $t<T_s<\phi(t)$, we have that $h(t)>0$ and $h(\phi(t))<0$, so $\phi'(t)<1$. 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 to

$$\begin{aligned}L(t,\phi(t)) &= \int_t^{T_s} h(z)\,dz+\int_{T_s}^{\phi(t)} h(z)\,dz\\ &= -\frac{(t-T_s)^2}{2}h'(\delta_t)+\frac{(\phi(t)-T_s)^2}{2}h'(\psi_t),\end{aligned}$$

where $\delta_t\in[t,T_s]$ and $\psi_t\in[T_s,\phi(t)]$. Since $L(t,\phi(t))=0$, and $\phi(t)>T_s$ and $t<T_s$,

solving this equation

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

which goes to -1 as t goes to T_s because both δ_t and ψ_t go to T_s and $h'(T_s) > 0$. It thus follows that $\phi'(T_s) = -1$. Now, keeping in mind that $h(T_s) = 0$, we can write:

$$\phi'(t) \;=\; \frac{h(t)}{h(\phi(t))} = \frac{h(t) - h(T_s)}{t - T_s}\frac{h(\phi(t)) - h(\phi(T_s))}{\phi(t) - \phi(T_s)}\frac{\phi(t) - \phi(T_s)}{t - T_s}.$$

Taking the limit $t \to T_s$, we obtain that $\phi'(t) \to \phi'(T_s)$.

II.2 Comparative Static

We consider $t < T_s$. Because $\phi(t) > T_s$, it follows that

$$\frac{\partial L}{\partial \phi} \;=\; \rho\,(s - \mu_h(\phi))\,e^{\rho\phi} < 0. \tag{II.2}$$

II.2.1 $\phi(t)$ is increasing in s

Also, taking partial derivatives wlith respect to s:

$$\frac{\partial L}{\partial s} \;=\; \int_t^{\phi} e^{\rho z}\,dz > 0,$$

implying, together with (II.2), by an application of the IFT, that $\phi(t)$ is increasing in s.

II.2.2 $\phi(t)$ is decreasing in ρ

Taking partial derivative with respect to ρ, evaluated $\phi(t)$

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

where the third line follows because, over $[t, T_s]$, $s - \mu_h(z)$ is positive so $z\,(s - \mu_h(z))$ is bounded above by $T_s\,(s - \mu_h(z))$. Over $[T_s, \phi(t)]$, $s - \mu_h(z)$ is negative so $z\,(s - \mu_h(z))$

is also bounded above by $T_s\left(s-\mu_h(z)\right)$. It thus follows that $\phi(t)$ is decreasing in ρ.

II.2.3 $\phi(t)$ is decreasing in $\mu_h(\cdot)$

Now suppose that $\mu_h(t)$ increases with some parameter θ. We then have:

$$\frac{\partial L}{\partial \theta} = -\int_t^{\phi} \frac{\partial \mu_h}{\partial \theta}(z) e^{\rho z}\, dz < 0,$$

so $\phi(t)$ decreases with θ.

II.2.4 $\phi(t)$ converges to T_s, uniformly in t

We start by extending Lemma 10:

Lemma 13 (Preliminary result)**.** *Let $f(t)$ be some bounded measurable function, continuous at $t=0$. Let $\{\psi_n\}$ be a positive sequence converging to zero, and ρ_n a sequence converging to infinity. Then, for every $t_{\max} < 0$,*

$$\int_t^{\delta_n} f(z)\rho e^{\rho_n z}\, dz \to f(0),$$

uniformly over all sequences $\{\delta_n\}$ such that $0 \le \delta_n \le \psi_n$, and all $t \in (-\infty, t_{\max}]$.

To see this, we calculate

$$\begin{aligned}
&\left|\int_t^{\delta_n} f(z)\rho_n e^{\rho_n z}\, dz - f(0)\right| = \left|\int_t^{\delta_n} \left(f(z)-f(0)\right)\rho_n e^{\rho_n z}\, dz - f(0)\left[1-e^{\rho_n\delta_n}+e^{\rho_n t}\right]\right| \\
< &\left|\int_t^{\delta_n} \left(f(z)-f(0)\right)\rho_n e^{\rho_n z}\, dz\right| + f(0)\left(e^{\rho_n t_{\max}} + e^{\rho_n\psi_n} - 1\right) \\
< &\left|\int_t^{\delta_n} \left(f(z)-f(0)\right)\rho e^{\rho_n z}\, dz\right| + o(1),
\end{aligned}$$

where, in the above and in what follows, $o(1)$ denotes a sequence of function converging to zero as n goes to infinity, uniformly over all sequences $0 \le \delta_n \le \psi_n$ and over $t \in (-\infty, t_{\max}]$. Now, because $f(t)$ is continuous at $t=0$, for every $\varepsilon > 0$ there is some $0 < \eta < t_{\max}$ such that $|f(t)-f(0)| < \varepsilon/2$ whenever $|t| < \eta$. Further, for n large enough, $\psi_n < \eta$ and therefore $\delta_n < \eta$. Thus, for n large enough, the last expression is bounded above by

$$
\begin{aligned}
& \left|\int_t^{-\eta}\left|f(z)-f(0)\right|\rho e^{\rho_n z}\,dz+\left|\int_{-\eta}^{\delta_n}\right|f(z)-f(0)\right|\rho e^{\rho_n z}\,dz+o(1) \\
< \;& 2\sup|f(z)|\left(e^{-\rho_n\eta}-e^{-\rho_n t}\right)+\frac{\varepsilon}{2}\left(e^{\rho_n\delta_n}-e^{-\rho_n\eta}\right)+o(1) \\
< \;& 2\sup|f(z)|e^{-\rho_n\eta}+\frac{\varepsilon}{2}e^{\rho_n\psi_n}+o(1),
\end{aligned}
$$

which is less than ε for n large enough, for sequences $0 \le \delta_n \le \psi_n$ and all times $t \in (-\infty, t_{\max}]$.

Now turning to the behavior of $\phi(t)$ as ρ goes to infinity, we first note that $\phi(t)$ is bounded below by T_s and is decreasing in ρ. So it has a limit $\phi^*(t)$, as ρ goes to infinity. Now note that

$$
\begin{aligned}
0 &= L(t,\phi(t))e^{-\rho\phi^*(t)} = \int_t^{\phi(t)}(s-\mu_h(z))\rho e^{\rho(z-\phi^*(t))}\,dz \\
&= \int_{t-\phi^*(t)}^{\phi(t)-\phi^*(t)}\Big(s-\mu_h(\phi^*(t)+z)\Big)\rho e^{\rho z}\,dz \\
&\to s-\mu_h(\phi^*(t)),
\end{aligned}
$$

by applying Lemma 13 with $f(z) = s - \mu_h(\phi^*(t) + z)$ and a lower bound of integration equal to $t - \phi^*(t) < t - T_s < 0$. It follows then that $\phi^*(t) = T_s$. The uniform convergence follows simply because $T_s \le \phi(t) \le \phi(0)$, and $\phi(0)$ converges to T_s.

III Investors Demographics and Order Flows

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 $\sigma \in \{hn, \ell n, ho, \ell o, hb, \ell b\}$, at time t, and we drop the time subscripts to simplify notations. The dynamics of distribution of are illustrated in Figure 3 and are summarized in the following ODEs:

$$\begin{aligned}
&\text{type } hn \qquad \dot{\mu}_{hn} = -\text{Mkt}_h + \text{LimExec}_h + \gamma\mu_{\ell n} && \text{(III.1)}\\
&\text{type } ho \qquad \dot{\mu}_{ho} = \text{Mkt}_h + \rho\mu_{hb} + \gamma\mu_{\ell o} && \text{(III.2)}\\
&\text{type } hb \qquad \dot{\mu}_{hb} = -\rho\mu_{hb} - \text{LimExec}_h + \gamma\mu_{\ell b} && \text{(III.3)}\\
&\text{type } \ell n \qquad \dot{\mu}_{\ell n} = \text{Mkt}_\ell + \text{LimExec}_\ell - \gamma\mu_{\ell n} && \text{(III.4)}\\
&\text{type } \ell o \qquad \dot{\mu}_{\ell o} = -\rho\mu_{\ell o} - \gamma\mu_{\ell o} && \text{(III.5)}\\
&\text{type } \ell b \qquad \dot{\mu}_{\ell b} = -\rho\mu_{\ell b} - \text{LimExec}_\ell + \text{LimSub} - \gamma\mu_{\ell b}, && \text{(III.6)}
\end{aligned}$$

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 (III.1), 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:

$$\text{LimSub} + \text{Mkt}_\ell = \rho\left(\mu_{\ell o} + \mu_{\ell b}\right). \tag{III.7}$$

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 (III.2), (III.3), (III.5) and (III.6), we obtain the natural condition:

$$\begin{aligned}
\text{Mkt}_h &= \left[\rho\mu_{\ell o} + \rho\mu_{\ell b} - \text{LimSub}\right] + \text{LimExec}_\ell + \text{LimExec}_h \\
&= \text{Mkt}_\ell + \text{LimExec}_\ell + \text{LimExec}_h
\end{aligned} \tag{III.8}$$

after plugging in equation (III.7). 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 (III.8), we obtain that $\text{Mkt}_\ell = \rho\mu_{hn}$. Next, plugging in (III.7), 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 of Proposition 2 for the flow of limit orders submitted during $[t, t + dt] \subseteq (0, T_s)$.

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 (III.8) 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 $t \geq T_s$. This confirm the formula of Proposition 3 for the flow of limit sell orders submitted during $[t, t+dt] \subseteq (T_s, T_f)$. Note that, by construction of T_f, at any time $t \in (T_s, T_f)$ there is a positive measure of limit sell orders in the book, so there is indeed enough limit orders to accommodate the net buy order flow $\rho\,(\mu_h - s)\; dt$.

The last thing to do is 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 $\text{Mkt}_h = \rho\mu_{\ell o}$.

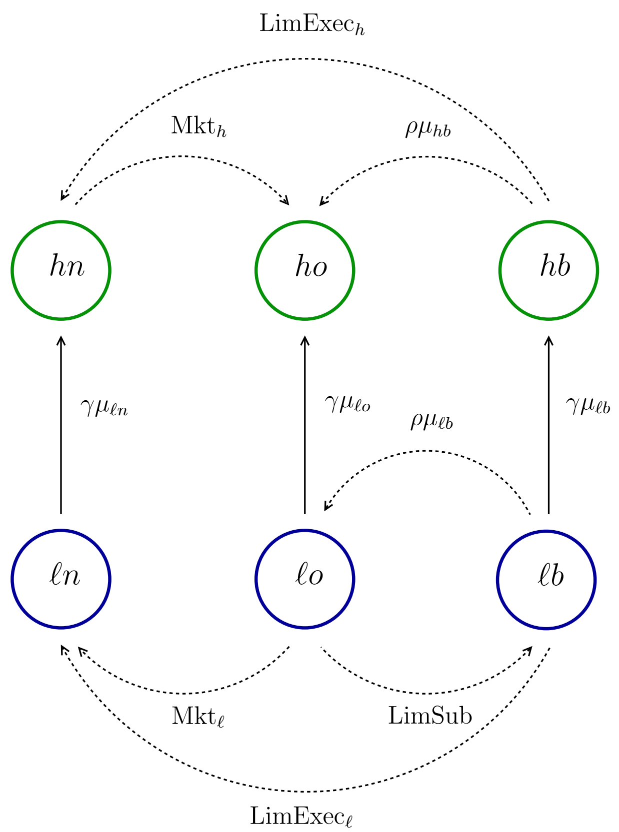

Figure 3: Inflows and outflows between types

IV A Two-population Model

This Addendum provides an extension of our model with two populations: a population of low-valuation investors who initially hold the asset, and a population of high-valuation investors who progressively enter the economy and purchase the asset. This specification is closer to the two period model of Grossman and Miller (1988), where sellers hold the asset in the first period and buyers exogenously enter the economy in the second period.

We assume that, at time zero, all assets are held by low valuation investors. As in our main model, these low valuation investors switch to a high utility with Poisson intensity, but with a parameter $\gamma(1-\varepsilon)$, where $\varepsilon \in [0,1]$. The measure of high-valuation investors at time t who previously were low-valuation investor is:

$$\mu_h^{\varepsilon}(t) = 1 - e^{-\gamma(1-\varepsilon)t}$$

Following the spirit of Grossman and Miller (1988), we also assume that there is an additional inflow of investors who progressively enter the economy without asset and with a high valuation. This may represent the arrival of additional capital, because each new entrant arrives in the market with a buying capacity of one share. We let the measure of "new entrant" at time t be some continuous and increasing function $\mu_h^{\text{new}}(t)$. Thus, the total measure of high-valuation investors in the economy is:

$$\mu_h(t) = \mu_h^{\varepsilon}(t) + \mu_h^{\text{new}}(t). \tag{IV.1}$$

The first term represents the measure of high-valuation investors who previously had a low valuation. The second term is the measure of "new entrant" high-valuation investors. Then, one easily shows:

Proposition 9 (Equilibrium with Two Populations)**.** *If $\varepsilon < 1$, then there is an MLOE that is identical to that of Theorems 1 and 2 after making the change of variable $\pi_h^{\varepsilon}(t,z) = 1 - e^{-\gamma(1-\varepsilon)(z-t)}$, and $\mu_h(t) = \mu_h^{\varepsilon}(t) + \mu_h^{new}$.*

When ε is very close to 1, it takes a very long time (on average) for low-valuation investors to recover, so most buy orders originate from "new entrant," just as envisioned by Grossman and Miller. What happens when $\varepsilon = 1$? Then one can show that the candidate of the Proposition remains an equilibrium. However, just as in Proposition 7, low-valuation investors would be indifferent regarding the execution time of their limit orders. Indeed, when $\varepsilon = 1$, low-valuation investors have no incentive to delay in order to mitigate the risk of being executed after they recover, since they never recover. As before, the proposition suggests that, in order to select among all these equilibria, it is enough to set ε arbitrarily close to 1 so that low-valuation investors strictly prefer to adopt decreasing limit order submission strategies.

IV.1 Proof of Proposition 9

In the candidate equilibrium, low and high valuation trading plans are those described in Lemma 3, Proposition 2, and Proposition 3. A limit sell order submitted at time t is executed at time $\phi(t)$, for the function $\phi(t)$ defined in Proposition 5, given the function $\mu_h(t) = \mu_h^\varepsilon(t) + \mu_h^{\text{new}}(t)$. Finally, the price is that of Proposition 6, given the function $\pi_h^\varepsilon(t, z)$. Given the functional form of $\pi_h^\varepsilon(t, z)$, the proof of Theorem 2 shows that the price path is indeed increasing.

To verify that this candidate is indeed an equilibrium, we proceed in two steps. First, we verify market clearing using the analysis of Addendum III, given the function $\mu_h(t) = \mu_h^\varepsilon(t) + \mu_h^{\text{new}}(t)$. Second, we verify optimality using the same proof as in the main model: indeed, one easily sees that this proof of optimality does not depend on the particular functional form for $\phi(t)$, it only depends on $\phi(t)$ being a strictly decreasing function.