Available online at www.sciencedirect.com
ScienceDirect
Journal of Economic Theory 144 (2009) 532–564
JOURNAL OF Economic Theory
www.elsevier.com/locate/jet

Portfolio choice and pricing in illiquid markets

Nicolae Gârleanu [a,b,c,1]

[a] *Haas School of Business, University of California, Berkeley, CA 94720-1900, United States*
[b] *CEPR, United Kingdom*
[c] *NBER, United States*

Received 16 November 2007; final version received 28 July 2008; accepted 29 July 2008

Available online 20 November 2008

Abstract

This paper studies portfolio choice and pricing in markets in which immediate trading may be impossible. It departs from the literature by removing restrictions on asset holdings, and finds that optimal positions depend significantly and naturally on liquidity: When expected future liquidity is high, agents take more extreme positions, given that they do not have to hold those positions for long when they become undesirable. Consequently, larger trades should be observed in markets with more frequent trading. Liquidity need not affect the price significantly, however, because liquidity has offsetting impacts on different agents' demands. This result highlights the importance of unrestricted portfolio choice. The paper draws parallels with the transaction-cost literature and clarifies the relationship between the price level and the realized trading frequency in this literature.

JEL classification: D53; D61; G11; G12

Keywords: Liquidity; Discount; Portfolio choice; Trading delays; Search; Transaction costs

The fact that in many financial markets completing a trade may require a significant amount of time has motivated an increasing amount of research in the past few years. The present paper contributes two-fold to the literature on this aspect of market illiquidity. First, it introduces the important dimension of portfolio choice in the equilibrium model and shows explicitly how the optimal choices depend on the liquidity level. Second, it shows that with no restriction on

E-mail address: garleanu@haas.berkeley.edu.
[1] Address for correspondence: Haas School of Business, University of California, Berkeley, CA 94720-1900, United States. Fax: +1 510 643 1420.

0022-0531/$ – see front matter
doi:10.1016/j.jet.2008.07.006

the portfolio choice, the equilibrium impact of illiquidity on prices is considerably smaller than derived in this literature. I highlight clearly the mechanism delivering this result and point out the features required to obtain, instead, a significant illiquidity price impact. In addition, I show that this mechanism also applies in a setting with transaction costs, and explain the relationship between the illiquidity discount and the observed trading frequency in such a setting.

Clear evidence exists that organizing a trade in financial markets may incur important delays. Even in the Federal Funds market, for instance, which is one of the most liquid over-the-counter markets, Ashcraft and Duffie [3] find extensive evidence of pervasive, albeit small, search frictions. In the corporate- and municipal-bond markets, finding an appropriate trade counterparty is considerably more difficult, consistent with the lack of incentives to specialize in small issues with low turnover.[2] Similarly, arranging a trade for a block of shares or complex derivatives is time consuming.[3] Investments that are essentially non-tradeable for certain periods of time, such as private equity, are even more illiquid.[4,5]

This paper studies optimal portfolio choice and equilibrium prices in this type of markets. Thus, it aims to answer such questions as: How large a stake in private equity should one take, given that it cannot be changed for a lengthy period of time? Given an investor's inability to change her corporate-bond position quickly, what price should she pay for a block of these bonds?

Delays to trading in financial markets have been recognized and studied by a variety of authors, such as Duffie et al. [14], Weill [34], Vayanos and Wang [32], and others. As a reasonable starting point, these models effectively limit the agents to two asset positions—low and high—and study price formation when trading can only occur following successful search. The present paper extends the literature by allowing, in a model similar to that of Duffie et al. [14], that agents choose the sizes of their investments with no restrictions, thus enabling the study of portfolio choice in this context. The results are interesting and intuitive: The less easily agents can trade in the future, the less extreme positions they take currently in order to avoid holding highly disadvantageous positions for extended periods at some later time. For instance, when expecting difficult future trading, an institution with current high value for a particular corporate bond—say, due to a low correlation with the rest of its portfolio—should buy a smaller amount of the bond than it would in a perfect market, anticipating a reduction in its value for the bond. The size of individual trades, therefore, is smaller in markets in which agents expect to be able to trade only infrequently in the future. Of course, since more difficult search means less frequent trades, and results in smaller trade sizes, it also reduces the volume of trade.

The second result concerns the equilibrium price, which, unlike portfolio choice, depends little on illiquidity.[6] The intuition for the mechanism at play is completely transparent: Lower liquidity reduces the demand of agents whose value for the asset is above its long-run average, and therefore expected to decline, but it increases the demand of agents with below-average

[2] In the average municipal bond issue, for instance, there is roughly one trade a month, as reported by Green et al. [19].

[3] For convertible bonds, a practitioner estimate of the monthly amount that one can sell before having to accept deep discounts is $150 million, but a number of investment funds sometimes hold, and occasionally want to trade quickly, many times that amount. See Mitchell et al. [27] for some examples.

[4] The minimum holding period before selling restricted securities publicly, as specified by Rule 144(d), is one year. Some securities are further restricted by IPO lockup-periods.

[5] Duffie et al. [14] provide other examples and references.

[6] This result is reminiscent of similar conclusions drawn when illiquidity is due to transaction costs. I discuss the relationship in more detail below and in Section 4.2.

value. The impact on the aggregate demand can consequently be quite small—in the present paper, the effect is literally zero in the approximation used to obtain a closed-form solution, and negligible when solving the model numerically for reasonable parameterizations. On the other hand, concurrently with a small price impact of illiquidity, a substantive impact on welfare can obtain, as the modeled gains from trade are quantitatively significant.

The result on the price effect appears at odds with the one in the search literature, which finds the price effect to be significant. I show that the difference is due to the portfolio constraints imposed in this literature, which keep traders from being marginal. Introducing binding short-sale constraints, for instance, in the present set-up results in a price that increases with liquidity; importantly, the effect can be quantitatively significant. The empirical support for the existing results of the search literature, therefore, is likely to owe in part to such trading restrictions as short-sale constraints or indivisibility that prevent (some) agents from adjusting their demand to the liquidity level.

More generally, the paper shows that, in order for illiquidity to have a significant price effect in the absence of binding portfolio constraints, one of the following conditions must hold; (i) the agents trading at a given point in time are not representative of all investors—maybe the ones keener to trade (say, sell) exert more effort and are therefore over-represented; or (ii) the slopes of the marginal utilities vary significantly with the asset holdings.[7] This observation constitutes a refinement of the current theoretical understanding of the interaction between liquidity and prices, and is of immediate practical use to researchers looking to build models of settings in which liquidity has an important price impact, as market observers argue and empirical studies suggest.

The paper relates to several bodies of research. First, I complement the literature on search in financial markets[8] by allowing agents to choose asset positions freely. I am consequently able to both characterize the optimal portfolio choice and add to the theory of the price impact of liquidity in this context.

Second, the paper relates to the more general literature on trading liquidity.[9] In this connection, it is worth starting with the observation that search frictions are different from both exogenous transaction costs and information frictions. To put it succinctly, the latter generate trading losses, exogenous in the case of transaction costs, whereas search frictions generate only costs of waiting, i.e., of *not trading*. These are utility losses determined endogenously. A consequence is that the current level of transaction costs or asymmetric information impacts the market price, whereas that of the search friction does not.[10] All of these frictions, however, result in asset misallocation and reduced trading volume, due to the anticipation of the costs.[11] Furthermore, the adjustments in trading diminish the impact of the friction on the price, as explained above.

To make the relationship with the transaction-cost literature clearer, I extend the model to incorporate transaction costs. I show explicitly that the mechanism involving the opposite effects

7 Lagos and Rocheteau [24] explores this avenue by specifying preferences directly over the asset holding.

8 In addition to the papers cited above, notable examples include Duffie et al. [12] and Vayanos and Weill [33] studying search as friction to shorting, Duffie et al. [13] and Weill [35] considering market-maker behavior, and Gârleanu and Pedersen [17] analyzing the equilibrium implications of risk management.

9 See Amihud et al. [2] for a recent survey of the literature.

10 Naturally, search frictions also have rich and unique implications for the dynamics of equilibria of economies outside steady state, but I do not explore this difference in the current paper. Section 4.2 is more detailed concerning specific differences between the implications of delayed trading and transaction costs.

11 See Gârleanu and Pedersen [16] for a simple illustration in the context of asymmetric information.

on different agents' demand schedules also generates the small magnitude of transaction-cost impacts on asset prices found in the theoretical literature (e.g., Constantinides [8], Vayanos [31], Huang [22], or Lo et al. [25][12]). However, the possibly large welfare impact obtaining shows that, in contrast to the conclusion of the transaction-cost literature, the small effect on price is not due to agents' utilities being relatively insensitive to their ability to trade. Also in contrast to existing results, such as those of Amihud and Mendelson [1], I find that the trading frequency has little or no direct impact on the price effect of transaction costs.[13] Forced exit, on the other hand, has the usual price-reduction effect reflecting the amortized future transaction costs. The distinction between trading frequency and (the reciprocal of) market-participation horizon can be of large importance in practice, and therefore of significance for empirical tests.

Finally, the issue of infrequent adjustment has also received attention recently. Following the seminal contribution of Grossman and Laroque [20], such papers as Gabaix and Laibson [15], Reis [28], and Chetty and Szeidl [7] model agents that adjust their consumption discretely. The focus of these papers, however, is the correlation between consumption and asset returns, rather than the price and portfolio impacts of trading frictions. Closer in spirit to the present paper is Longstaff [26], which uses numerical techniques to calculate the effects of a "blackout" period, during which one asset cannot be traded.

The paper is organized as follows. Section 1 presents the model and Section 2 defines and characterizes equilibrium in the economy under study. Section 3 discusses the main properties of the equilibrium, and provides a calibration. Section 4 addresses the relationships with the search and transaction-cost literatures. It also presents two extensions to the basic model that, in addition to being of independent interest, are helpful in understanding these relationships. Section 5 concludes.

1. Basic model

This paper considers a two-asset economy. One asset is riskless, pays interest at an exogenously given constant rate r, and is available in perfectly elastic supply. The other asset pays a cumulative dividend with i.i.d. Gaussian increments:

$$dD(t) = m_D\,dt + \sigma_D\,dB(t). \tag{1}$$

Here, m_D and σ_D are constants, and B is a standard Brownian motion with respect to the given probability space and filtration $(\mathcal{F}_t)$. The per-capita supply of this asset is Θ, and its price is determined in equilibrium.

There are a continuum of agents, with total mass normalized to 1. Agent i has a cumulative endowment process η^i, with

$$d\eta^i(t) = m_\eta\,dt + \sigma_\eta\,dB^i(t), \tag{2}$$

where the standard Brownian motion B^i is defined by

$$dB^i(t) = \rho^i(t)\,dB(t) + \sqrt{1-\rho^i(t)^2}\,dZ^i(t). \tag{3}$$

[12] Lo et al. [25] can also generate large effects. I discuss this in Section 4.2.4.

[13] Vayanos [31] shows that, while the impact owing purely to the trading frequency could be substantial, the overall impact of transaction costs is quite low, implying that the former is largely canceled by the adjustment in the risk agents take. I add to this conclusion by showing explicitly why the cancellation happens and the observed trading frequency is unrelated to the price discount.

The standard Brownian motion Z^i is independent of B, and $\rho^i(t)$ is the instantaneous correlation between the asset dividend and the endowment of agent i. I assume that ρ^i, referred to as the type of agent i, follows a Markov process on a finite state space with $J > 1$ points $1 \geqslant \rho_1 > \cdots > \rho_J \geqslant -1$. The transition intensity from state j to state l is denoted by α_{jl}. For simplicity, I assume that the Markov chain is irreducible.

Agents have von Neuman–Morgenstern utilities with constant-absolute-risk-aversion (CARA) felicity functions with coefficient $\gamma > 0$ and time preference at rate β. Changes in correlation between dividends and endowment induce them to want to trade. I assume, however, that they may be unable to trade immediately.

This kind of illiquidity is observed in numerous real-world settings. First, and most severe,[14] trading in some assets is exogenously prohibited for a period of time. A clear example is provided by restricted securities as defined by Rule 144(a), which include issues by private companies as well as private issues by public firms, such as private investments in public equity (PIPEs).[15] These securities cannot be traded publicly for at least 1 year since issuance, while the market for private trading is extremely thin. Similarly, many investment institutions, such as hedge funds, impose lock-up periods on investors' funds. Second, while legally not prohibited, trading some assets, such as emerging-market securities, may be impossible for important periods of time due to market closures or insufficient activity. Third, many assets are not traded in centralized markets. Here, in order to trade, an agent may have to search for a qualified counterparty, or an opportunity to trade. For instance, there are many assets, such as given corporate bonds, shares in companies emerging from Chapter 11, or real-estate investments, that are only traded by a relatively small number of market participants, who have the required expertise. Finding such a participant that is able to take on a larger position, or willing to sell her stake, takes time. Additional time might also be necessary to convince the counterparty that the sale is not motivated by information. As noted in the introduction, even the most liquid over-the-counter markets are subject to delays—see, for instance, Ashcraft and Duffie [3] and the discussion in Vayanos and Weill [33].

I model the infrequent-trading feature by assuming that each agent can trade only at a subset of the time line. This subset can be random or deterministic, thus being able to capture scheduled closures and search difficulties. I describe such a general formulation, as well as the possibility of other parameters depending on time, in Appendix A, but, for simplicity,[16] I shall focus on the assumption that each agent trades at Poisson arrival times. More specifically, I assume that each agent comes across a trading post (or competitive marketmaker), where she takes the price as given, at the arrival times of a Poisson process with constant intensity $\lambda \in (0, \infty)$. These Poisson processes, B, Z^i, and ρ^i for all i are mutually independent. The law of large numbers will be assumed to hold throughout.[17]

[14] Arguably, Longstaff [26] gives what may be even more extreme examples of untradeable assets, such as human capital.

[15] In the case of PIPEs, an investor's inability to trade out of a position depends on the ease with which she can short public shares in the firm. Tannenbaum and Pinedo [30] discuss contractual provisions that prohibit investors to engage in hedging activities. Furthermore, they also report that the SEC has ruled, in its *Telephone Interpretations*, that an investor may not short before the registration of the new securities is effective, which is typically expected to take at least 30 to 120 days (see Klein [23]). Finally, Chaplinsky and Haushalter [6] argue that, due to the specifics of the shorting market in the underlying stocks, it is usually difficult to establish and maintain short positions.

[16] More specifically, in order to obtain a stationary, rather than a time-varying, solution.

[17] The results in Duffie and Sun [11] ensure the existence of a discrete-time version of the model where the law of large numbers holds.

An agent possessing θ shares of the asset has a value function defined as

$$V(w,\rho,\theta)=\sup_{\bar{c},\bar{\theta}}\mathrm{E}_t\left[-\int_t^{\infty}e^{-\beta(s-t)}e^{-\gamma\bar{c}_s}\,ds\;\middle|\;\rho(t)=\rho,\ W_t=w,\ \bar{\theta}(t)=\theta\right], \tag{4}$$

s.t.

$$dW_t=(rW_t-\bar{c}_t)\,dt+\bar{\theta}(t)\,dD_t+d\eta_t-P_t\,d\theta_t, \tag{5}$$

where W is the agent's total cash holding at any point in time, $\bar{c}$ the agent's consumption, and $\bar{\theta}$ the number of shares he owns in the risky asset. The optimization problem is further constrained by the requirement that the asset holding be chosen only at the arrival times of the Poisson process. To avoid Ponzi schemes, I impose the transversality condition

$$\lim_{T\to\infty}e^{-\beta(T-t)}E_t\left[e^{-r\gamma W_T}\right]=0 \tag{6}$$

and require $|\theta|<M$ for some sufficiently large, but otherwise arbitrary, $M>0$.

2. Equilibrium definition and characterization

An equilibrium in this setting is defined in the usual way. I concentrate on stationary and wealth-independent[18] equilibria, where all agents of type ρ_j trade to the same position θ_j, independent of their wealth levels, at any point in time. As agents need not be able to trade instantly in response to type changes, some agents may not hold their optimal positions. Let μ_{ij} denote the mass of agents of type ρ_i holding position θ_j. It is easy to see that, with $i\neq j$, the rates of change of μ_{ij} and μ_{jj} are given by

$$\dot{\mu}_{ij}=-\mu_{ij}\sum_{k\neq i}\alpha_{ik}+\sum_{k\neq i}\alpha_{ki}\mu_{kj}-\lambda\mu_{ij}, \tag{7}$$

$$\dot{\mu}_{jj}=-\mu_{jj}\sum_{k\neq j}\alpha_{jk}+\sum_{k\neq j}\alpha_{kj}\mu_{kj}+\lambda\sum_{k\neq j}\mu_{jk}. \tag{8}$$

Stationarity requires that these rates of change be zero.

Definition 1. A stationary equilibrium consists of a set of masses μ_{ij}, $i,j=1,\dots,J$, a set of positions θ_j, $j=1,\dots,J$, a function $c(w,\theta,\rho)$, and a price P such that

1. Agents optimize: $c(w,\theta,\rho)$ attains the maximum in (4) and θ_j attains the maximum of $V(w-P\theta,\rho_j,\theta)$.
2. All agent groups have constant masses: (7) and (8) hold with 0 on the left-hand side.
3. The asset market clears: $\sum_j\sum_k\mu_{jk}\theta_k=\Theta$.

Following Duffie et al. [14], hereafter DGP,[19] I conjecture a value function of the form

$$V(w,\rho,\theta,t)=-e^{-r\gamma(w+\bar{a}+a(\theta,\rho))}, \tag{9}$$

[18] This feature is due to all agents having CARA utility.

[19] In order to be self-contained, I give a sketch of the argument in Appendix A.

where

$$\bar{a} = \frac{1}{r}\left(\frac{\log r}{\gamma} + m_\eta - \frac{1}{2}r\gamma\sigma_\eta^2 - \frac{r-\beta}{r\gamma}\right) \tag{10}$$

is a constant.

Let $a_j(\theta) = a(\theta, \rho_j)$ be the value-function coefficient for an agent of type ρ_j. These coefficients obey a set of Hamilton–Jacobi–Bellman equations that, after dividing across by $e^{-r\gamma(w+a_j(\theta))}$, simplify to

$$\begin{aligned} ra_j(\theta) = \kappa(\theta, \rho_j) - \sum_l \alpha_{jl} \frac{e^{-r\gamma(a_l(\theta)-a_j(\theta))} - 1}{r\gamma} \\ - \lambda \sup_{\bar{\theta}} \frac{e^{-r\gamma(-P(\bar{\theta}-\theta)+a_j(\bar{\theta})-a_j(\theta))} - 1}{r\gamma}, \end{aligned} \tag{11}$$

where

$$\kappa(\theta, \rho) = \theta m_D - \frac{1}{2}r\gamma\left(\theta^2\sigma_D^2 + 2\rho\theta\sigma_D\sigma_\eta\right) \tag{12}$$

is the (mean-variance) instantaneous benefit to the agent from holding position θ when of type ρ.

In a stationary equilibrium, all agents of the same type ρ_j choose the same position θ_j. The positions are determined so that agents maximize their utilities, implying that, at $\theta = \theta_j$,

$$0 = \frac{d}{d\theta} V(w - P\theta, \rho_j, \theta), \tag{13}$$

which simplifies to

$$P = a_j'(\theta_j). \tag{14}$$

Differentiating (11) with respect to θ yields

$$\begin{aligned} ra_j'(\theta) = \sum_l \alpha_{jl} e^{-r\gamma(a_l(\theta)-a_j(\theta))}\left(a_l'(\theta) - a_j'(\theta)\right) \\ + \lambda e^{-r\gamma(-P(\theta_j-\theta)+a_j(\theta_j)-a_j(\theta))}\left(P - a_j'(\theta)\right) + \kappa_1(\theta, \rho_j), \end{aligned} \tag{15}$$

where κ_1 is the partial derivative of κ with respect to its first argument.

Eqs. (11)–(15) cannot be solved in closed form. Consequently, I resort to an approximation that ignores terms of order higher than 1 in $(a_l(\theta) - a_j(\theta))$—that is, I use $\frac{e^{-r\gamma x}-1}{r\gamma} \approx x$ in (11). The accuracy of this approximation depends on the size of $r\gamma(a_l(\theta) - a_j(\theta))^2$, which can be shown to be small when $r^3\gamma^3(\rho_1 - \rho_J)^2\sigma_D^2\sigma_\eta^2$ is small. To make precise these statements, I follow Vayanos and Weill [33] and consider the limit as $\gamma \to 0$ while holding $\gamma\sigma_D^2$ and $\gamma\sigma_\eta^2$ constant. In effect, this maintains risk aversion towards dividend and endowment flows, while inducing risk neutrality towards changes in type and arrival of trading opportunities. A rigorous statement is made in Theorem 1 below. The numerical example in Section 3.4 demonstrates that the approximation is accurate for reasonable parameters.

The approximation yields

$$ra_j(\theta) = \sum_l \alpha_{jl}\left(a_l(\theta) - a_j(\theta)\right) + \lambda\left(-P(\theta_j - \theta) + a_j(\theta_j) - a_j(\theta)\right) + \kappa(\theta, \rho_j) \tag{16}$$

and

$$ra'_j(\theta) = \sum_l \alpha_{jl}\big(a'_l(\theta) - a'_j(\theta)\big) + \lambda\big(P - a'_j(\theta)\big) + \kappa_1(\theta, \rho_j). \tag{17}$$

Note that the approximate HJB equations (16) obtain exactly when agents are risk-neutral, but the benefit from holding the asset is quadratic. More precisely, they obtain when the value functions are given by

$$\begin{aligned} a_j(\theta) = \sup_{\bar\theta} \mathrm{E}_t\Bigg[& \int_t^\infty e^{-r(s-t)}\kappa\big(\bar\theta(s), \rho(s)\big)\, ds \\ & - \sum_{s=t}^\infty e^{-r(s-t)} P_s \Delta\bar\theta(s) \,\Big|\, \rho(t) = \rho(0),\ \bar\theta(t) = \theta \Bigg], \end{aligned} \tag{18}$$

where trading is only possible at the arrival times of the individual Poisson process.

An immediate consequence is that, in equilibrium, for all $k = 1, \dots, J$ it holds approximately[20] that

$$P = \mathrm{E}_t\left[\int_t^\infty e^{-r(s-t)}\kappa_1\big(\theta(s), \rho(s)\big)\, ds \,\Big|\, \theta(t) = \theta_k,\ \rho(t) = \rho_k \right]. \tag{19}$$

Eq. (19) is intuitive, stating that the price equals the sum of the stream of discounted marginal utilities from the asset at all future times. (The equation is easily derived by considering permanent deviations in holdings from the optimal ones.)

Eq. (19) holds for all agents trading at a given time. I could add the left- and right-hand sides over these agents to obtain the price, but instead, in the interest of transparency, I follow the route of writing out explicitly the demand of an agent given the opportunity to trade. Specifically, rewrite (19) as

$$P = \mathrm{E}_t\left[\int_t^\tau e^{-r(s-t)}\kappa_1\big(\theta_k, \rho(s)\big)\, ds \,\Big|\, \rho(t) = \rho_k \right] + \mathrm{E}_t\big[e^{-r(\tau-t)}\big]P \tag{20}$$

and define

$$\begin{aligned} \tilde\mu_{jk} &= \left(\mathrm{E}_t\left[\int_t^\tau e^{-r(s-t)}\, ds \right] \right)^{-1} \mathrm{E}_t\left[\int_t^\tau e^{-r(s-t)} 1_{(\rho(s)=\rho_j)}\, ds \,\Big|\, \rho(t) = \rho_k \right] \\ &= r\big(1 - \mathrm{E}_t\big[e^{-r(\tau-t)}\big]\big)^{-1} \mathrm{E}_t\left[\int_t^\tau e^{-r(s-t)} 1_{(\rho(s)=\rho_j)}\, ds \,\Big|\, \rho(t) = \rho_k \right], \end{aligned} \tag{21}$$

where τ is the first arrival time of a trading opportunity after time t. Thus, the quantities $\tilde\mu_{jk}$ give, for all j, the relative payoff weights of the promises to receive a unit flow of consumption at any future time s such that $\rho(s) = \rho_j$, as long as τ has not occurred by s, given that $\rho(t) = \rho_k$. The quantities $\tilde\mu$ are easily computed using standard Markov-chain calculations.

Given the linearity of κ_1 in ρ_j, Eq. (20) is therefore rewritten as

[20] Throughout the paper this term will be used to mean “in the limit sense of Theorem 1.”

$$P = \frac{1}{r}\kappa_1\left(\theta_k, \sum_j \tilde{\mu}_{jk}\rho_j\right), \tag{22}$$

which is easily solved for the optimal quantity choice θ_k as a function of P:

$$\theta_k = \frac{1}{\gamma\sigma_D^2}\left(\frac{m_D}{r} - P\right) - \frac{\sigma_\eta}{\sigma_D}\sum_j \tilde{\mu}_{jk}\rho_j. \tag{23}$$

As is natural, these demand schedules[21] depend on the liquidity level. Importantly, some demands increase, while others decrease with liquidity—precise statements are made in Section 3.1 below—which leads to the price result I now derive.

Let $\mu_{\cdot k}$ denote the mass of all agents holding θ_k and $\mu_{k\cdot}$ the mass of all agents of type k. Note that, since the net flux into the group holding θ_k is $\lambda(\mu_{k\cdot} - \mu_{\cdot k}) = 0$ per unit of time, $\mu_{k\cdot} = \mu_{\cdot k} \equiv \mu_k$. Note further that

$$\begin{aligned}\sum_k \mu_k \mathrm{E}_t\left[1_{(\rho(s)=\rho_j)} \mid \rho(t) = \rho_k\right] &= \sum_k \mathrm{E}_t\left[1_{(\rho(s)=\rho_j, \rho(t)=\rho_k)}\right] \\ &= \mu_j.\end{aligned} \tag{24}$$

Consequently, multiplying Eq. (23) by μ_k and adding over k yields

$$\begin{aligned}\Theta &= \frac{1}{\gamma\sigma_D^2}\left(\frac{m_D}{r} - P\right) - \frac{\sigma_\eta}{\sigma_D}\sum_{j,k} \tilde{\mu}_{jk}\rho_j \\ &= \frac{1}{\gamma\sigma_D^2}\left(\frac{m_D}{r} - P\right) - \frac{\sigma_\eta}{\sigma_D}\bar{\rho},\end{aligned}$$

so that

$$P = \frac{m_D}{r} - \gamma\left(\Theta\sigma_D^2 + \sigma_D\sigma_\eta\bar{\rho}\right) \equiv P^W. \tag{25}$$

The term $\bar{\rho}$ denotes the average correlation in the economy and is independent of the level of liquidity λ. Consequently, so is P, which equals its value in a Walrasian market,[22] P^W. Using this expression,

$$\theta_k = \Theta + \frac{\sigma_\eta}{\sigma_D}\left(\bar{\rho} - \sum_j \tilde{\mu}_{jk}\rho_j\right). \tag{26}$$

The Walrasian holdings can be obtained in the limit as $\lambda \to \infty$, which gives $\tilde{\mu}_{jk} \to 1_{(j=k)}$, thus implying

$$\theta_k^W = \Theta + \frac{\sigma_\eta}{\sigma_D}(\bar{\rho} - \rho_k). \tag{27}$$

The results derived above are collected in the following theorem, and discussed in the next section. All required proofs are provided in Appendix A.

[21] Note that this equation gives the demand of an agent as a function of a price that will remain constant indefinitely. The current price, to be paid when purchasing, can be separated from the future, resale price, to obtain the demand schedule in the usual sense, but that will not be necessary.

[22] As is the case with the holdings, derived below, this is the exact price for any value of γ. See Theorem 1.

Theorem 1. *The economy studied has a stationary equilibrium, determined by Eqs.* (11), (12)*, and* (14)*. In this equilibrium, the value function and consumption are given by*

$$V(w,\rho,\theta) = -e^{-r\gamma(w+\bar{a}+a(\rho,\theta))},$$
$$c(w,\rho,\theta) = -\frac{\log(r)}{\gamma} + r\big(w+\bar{a}+a(\rho,\theta)\big).$$

Furthermore, fix parameters $\bar{\gamma}$, $\bar{\sigma}_D$, and $\bar{\sigma}_\eta$ and let $\sigma_D = \bar{\sigma}_D\sqrt{\bar{\gamma}/\gamma}$ and $\sigma_\eta = \bar{\sigma}_\eta\sqrt{\bar{\gamma}/\gamma}$. Then, as γ goes to zero, the limit price is

$$P = \frac{m_D}{r} - \bar{\gamma}\big(\bar{\sigma}_D^2\Theta + \bar{\rho}\bar{\sigma}_D\bar{\sigma}_\eta\big) \tag{28}$$

while the limit positions equal

$$\theta_k = \Theta + \frac{\bar{\sigma}_\eta}{\bar{\sigma}_D}\left(\bar{\rho} - \sum_j \tilde{\mu}_{jk}\rho_j\right). \tag{29}$$

Finally, in the Walrasian economy, the price and quantities are given exactly by

$$P^W = \frac{m_D}{r} - \bar{\gamma}\big(\bar{\sigma}_D^2\Theta + \bar{\rho}\bar{\sigma}_D\bar{\sigma}_\eta\big), \tag{30}$$

$$\theta_k^W = \Theta + \frac{\bar{\sigma}_\eta}{\bar{\sigma}_D}(\bar{\rho} - \rho_k). \tag{31}$$

3. Equilibrium properties

In this section I discuss the main properties of the equilibrium—the asset holdings and the price—as well as a notion of welfare that I introduce below. I supplement the discussion with a calibrated example that demonstrates the accuracy of the approximation for empirically relevant parameters and thus lends support to the conclusions drawn based on this approximation.

3.1. Demand schedules and holdings

Equilibrium holdings, given by (29), and demand schedules, given by (23), depend in the same way on liquidity and therefore have the same intuitive properties. To keep the discussion simple, I focus on the holdings. The first term in (29) is the per-capita supply. The second reflects the difference in vulnerability to the asset-payoff risk between the average agent and the agent considered. Thus, if the correlation between the agent's endowment and the asset dividend is going to be relatively high, in expectation, until the next trading opportunity, then the agent will hold a lower position, and vice versa. In particular, if the agent can trade continuously, then the holding depends on the difference between the average correlation and her current correlation, as Eq. (31) shows.

The fact that agents tilt their portfolio toward the ones desired in likely future states suggests that they would take less extreme positions in illiquid markets: they want to avoid getting stuck with highly disadvantageous positions if they know that trading away is difficult. As a consequence of the less extreme positions, the average trade size would be smaller when the market is illiquid, thus reducing volume beyond the direct effect of a worse ability to conduct a trade.

Although intuitive, the notion that all agents' holdings (and demands) are less extreme in illiquid markets, due to the cost of having to maintain undesirable positions later, is not always true. As Proposition 2 shows, restrictions on the transition matrix of the correlation process are necessary. The reason is that an agent with, say, a relatively high valuation for the asset—low asset-endowment correlation—may have an even higher expected valuation in the near future, although her expected valuation in the long run is lower. Consequently, with lower liquidity, the agent is more concerned about the higher future valuation and therefore takes an even higher position, thus farther from the per-capita supply.[23] The result in part (i) of Proposition 2, which shows that, for high liquidity, the quantity that matters is the expected correlation conditional on a correlation change, $(\sum_{j\neq k}\alpha_{kj})^{-1}\sum_{j\neq k}\alpha_{kj}\rho_j$, makes this intuition precise. This conditional expected correlation can be lower than ρ_j for some j even if $\rho_j < \bar{\rho}$. Part (ii) of the proposition excludes this scenario by requiring every agent's valuation to be mean-reverting.

The following holds.[24]

Proposition 2. (i) *For any trading frequency $\lambda < \infty$, $\theta_1^W < \theta_k < \theta_J^W$ for all k. There exists $\underline{\lambda} < \infty$ such that, for $\lambda > \underline{\lambda}$, θ_k is monotonic in λ for all k. Furthermore, θ_k increases strictly in λ for $\lambda > \underline{\lambda}$ if and only if*

$$\frac{\sum_{j\neq k}\alpha_{kj}\rho_j}{\sum_{j\neq k}\alpha_{kj}} > \rho_k, \tag{32}$$

and vice versa. In particular, θ_1 is decreasing and θ_J is increasing in λ for $\lambda > \underline{\lambda}$.

(ii) *For all k, if $E_0[\rho(t) \mid \rho(0) = \rho_k]$ is monotonic in t then θ_k is monotonic and $|\theta_k - \Theta|$ increasing in t.*[25]

(iii) *If there are only two types ($J = 2$), then*

$$\begin{aligned}
\theta_1 &= \Theta - \alpha_{12}\left(\frac{1}{\alpha_{12}+\alpha_{21}} - \frac{1}{r+\lambda+\alpha_{12}+\alpha_{21}}\right)\frac{\sigma_\eta(\rho_1-\rho_2)}{\sigma_D} \\
&= \theta_1^W + \frac{\alpha_{12}}{r+\lambda+\alpha_{12}+\alpha_{21}}\frac{\sigma_\eta(\rho_1-\rho_2)}{\sigma_D}, \\
\theta_2 &= \Theta + \alpha_{21}\left(\frac{1}{\alpha_{12}+\alpha_{21}} - \frac{1}{r+\lambda+\alpha_{12}+\alpha_{21}}\right)\frac{\sigma_\eta(\rho_1-\rho_2)}{\sigma_D} \\
&= \theta_2^W - \frac{\alpha_{21}}{r+\lambda+\alpha_{12}+\alpha_{21}}\frac{\sigma_\eta(\rho_1-\rho_2)}{\sigma_D},
\end{aligned}$$

and θ_1 and θ_2 are monotonically decreasing, respectively increasing, in λ for all λ. The trade size, namely $\theta_2 - \theta_1$, the rate with which agents trade, namely

$$\lambda(\mu_{12}+\mu_{21}) = 2\lambda\frac{\alpha_{12}\alpha_{21}}{(\alpha_{12}+\alpha_{21})(\lambda+\alpha_{12}+\alpha_{21})},$$

and the trading volume, namely

$$\frac{1}{2}\lambda(\mu_{12}+\mu_{21})(\theta_2-\theta_1),$$

all increase with λ.

---

[23] Since the agent has a relatively high need for the asset, she takes a larger position than the per-capita supply.

[24] From now on, I restrict attention to the approximation, meaning that the precise formulation of all statements involves letting $\gamma \to 0$ with σ_D and σ_η as in the second part of Theorem 1.

[25] The condition holds, for instance, if α_{ij} is independent of i for all j. See Lagos and Rocheteau [24] for a related result.

The result in part (iii) on trade characteristics (trade size and volume) helps point out the complex impact of liquidity on trading volume: past liquidity determines the number of agents ($\mu_{12} + \mu_{21}$) that would trade if given the opportunity (this decreases with the level of liquidity), current liquidity determines the rate (λ) with which such agents actually get to trade, while future liquidity determines the positions to which they wish to trade, thus influencing the average trade size $\theta_2 - \theta_1$.[26]

3.2. Price

I now turn to the equilibrium price, also stated in Theorem 1. The main result in this connection is that the price level is the same as in a Walrasian market—in particular, it is independent of the liquidity level. This conclusion may be surprising, as it appears to run counter to the intuition that illiquidity decreases an asset's price, an intuition formalized, for instance, in DGP and Weill [34].

One may try understanding the result in at least two ways. The technical one rests on two elements: the flow of marginal utility is linear in asset holding and agent type, and agents trading today are representative of the entire population at all times. Consequently, when aggregating the marginal utilities of the trading agents, the marginal utility of the representative agent is obtained regardless of the liquidity level.

However, a valuable general intuition is also highlighted here: the demand (given in Eq. (23)) increases with liquidity for some agents—typically, agents whose correlation is expected to increase[27]—and decreases for others. Aggregate demand, therefore, need not be affected significantly. In other words, the market may clear without large price movements—certainly with smaller movements than if one side of the market were constrained with respect to its holdings so as to render its demand price inelastic, as in the search literature.

In Section 4 I discuss the relationship with this literature, including the different conclusion regarding price sensitivity to liquidity, as well as the relationship with the transaction-cost literature. I also present a couple of extensions to the model that facilitate this discussion.

3.3. Welfare

I provide here a brief analysis of welfare in this model. It is intuitive that a higher liquidity level λ should translate into a higher welfare, as it enables a better allocation of the dividend risk at no cost. Formalizing the welfare effect and contrasting it with the price effect is nevertheless instructive. When considering welfare, the literature—starting with Constantinides [8]—concludes that some forms of illiquidity have little effect on welfare, and therefore on prices as well. This section and the numerical example following show that the small price effect can obtain in conjunction with a significant welfare effect.

To make meaningful welfare comparisons among economies with different liquidity levels λ, I move away from steady state and start each economy at time 0 with a Walrasian market where

[26] This point can be made even more saliently in a model that departs from steady-state analysis to allow for liquidity to change at some time T from λ to a different level, λ'. It then follows, under natural conditions, that $\theta_2(t) - \theta_1(t)$ increases with λ' for $t \leqslant T$, whereas $\mu_{12}(t) + \mu_{21}(t)$ decreases with λ for $t \geqslant T$. See Proposition 5 in Appendix A.

[27] See Proposition 2(i) above for a precise statement.

all agents participate, after which trading is governed by the search-type illiquidity modeled above.[28]

One measure of an agent's utility is the certainty equivalent $a(\theta, \rho) + \bar{a}$, which reflects the agent's average consumption and the total risk faced throughout his life-time. As a measure of welfare, I use the sum of all agents' certainty equivalents at time 0. The advantage of using certainty equivalents is that the resulting measure is invariant to wealth transfers. The summing over all agents' certainty equivalents can be thought of as taking the expectation of an agent's certainty equivalent just before the revelation of his type at time 0.

The modification to the model does not affect its approximate solution.[29] Time-0 welfare can be written as

$$\mathcal{W} = \bar{a} + \sum_k \mu_k \big(a(\theta_k, \rho_k) - P\theta_k\big) + P\Theta, \tag{33}$$

where $a(\theta_k, \rho_k)$ solves the system of Eqs. (16). Note that the masses μ_k and P are independent of λ, whereas $a(\theta_k, \rho_k) - P\theta_k$ increases with λ. The latter statement follows from the optimality of the trading strategy of the agent ρ_k: he could choose to trade only with a probability $\lambda'/\lambda < 1$ whenever given the possibility, and change his portfolio accordingly, but does not.

The welfare as defined here captures exclusively the hedging benefits from being able to participate in such a market. These benefits increase with liquidity, even if the price is (approximately) constant. The calibrated example below demonstrates the quantitative relevance of this statement.

3.4. A calibrated example

To illustrate the theoretical results derived so far, as well as the appropriateness of the approximation considered, I calibrate the model. Given the difficulty of pinpointing parameters describing markets for which the kind of illiquidity studied here is relevant, the results below should be viewed mainly as suggestive. Extensive sensitivity analyses, however, show the qualitative properties of these results to be robust.

As a base-case specification, I set a number of parameters following Lo et al. [25]—in turn, these parameters are based on the stock-market estimation of Campbell and Kyle [5]. In addition to reflecting econometrically estimated properties of a specific market, this choice enables comparison with Lo et al. [25]. I discuss this paper and make the comparison in Section 4.2.

Thus, I start with $r = 0.037$, $\mu_D = 0.05$, $\sigma_D = 0.285$, and $P^W = 0.741$. I also let $\Theta = 1$. These values imply an annual dividend yield $\mu_D/P^W = 0.0675$, and therefore an equity return premium $\mu_D/P^W - r = 0.0305$, and a volatility $\sigma_D/P^W = 0.385$.[30] In addition, I consider two types of agents, and I set $\rho_2 = -0.25$, $\rho_1 = 0.75$, $\alpha_{21} = 0.5$, and $\alpha_{12} = 1.5$. These choices imply that the aggregate endowment contains no traded risk, i.e., $\bar{\rho} = 0$. Furthermore, for trading

[28] An alternative would be to distribute the asset in equal shares to all agents at time 0. The conclusion is identical.

[29] The only subtlety is that, since $\mu_{\cdot k}(0) = \mu_{k\cdot}(0)$ and $\frac{d}{dt}(\mu_{\cdot k}(t) - \mu_{k\cdot}(t)) = -\lambda(\mu_{\cdot k}(t) - \mu_{k\cdot}(t))$, it is still the case that $\mu_{\cdot k} = \mu_{k\cdot}$.

[30] Results are similar if the parameters are chosen to match yield spreads and volatilities from the corporate bond market. For instance, the average excess return on US corporate bonds has been estimated to be in the range 61bp–88bp, depending on the authors (see, for instance, de Jong and Driessen [10]), while the average yearly return volatility is around 6–8% (see Bao and Pan [4]). Over this range, with m_D so that the bond is priced at par in a liquid market and preserving the ratio σ_η/σ_D and the other parameters, the highest liquidity premium at $\lambda = 1$ is about 4bp and less than 1bp at $\lambda = 12$.

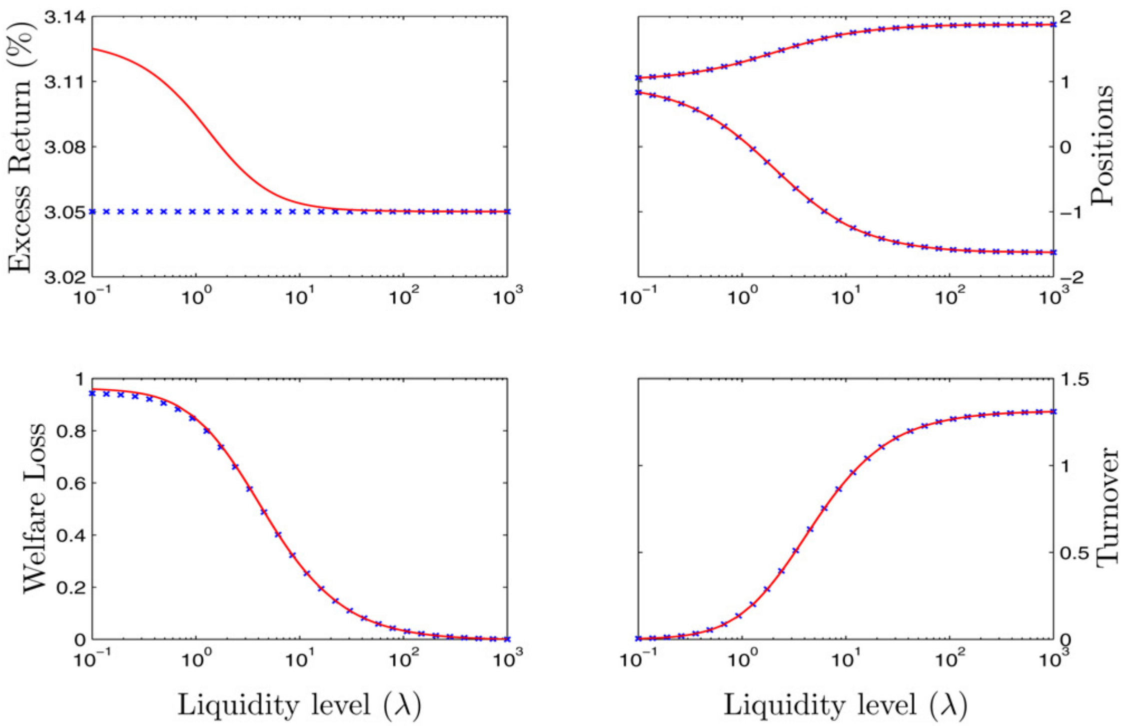

Fig. 1. Impact of illiquidity on the excess return (top left panel), holdings (top right panel), welfare (bottom left panel), and turnover (bottom right panel). Table 1 gives parameters. The continuous line plots the exact quantities computed numerically, whereas the line made of 'x' marks plots the quantities obtaining in the approximation.

Table 1
Parameters used to illustrate price and position impact of variable liquidity level. See Section 3.4 for a detailed explanation.

α_{12}	α_{21}	γ	ρ_1	ρ_2	r	m_D	σ_D	σ_η	Θ
1.5	0.5	7.52	0.75	−0.25	0.037	0.05	0.285	1	1

intensities that are not too low (e.g., $\lambda \geqslant 10$), a buyer's holding period is approximately $\alpha_{21}^{-1} = 2$ years, while for the seller it is about 8 months. Finally, fixing the difference $\rho_1 - \rho_2$, the extent of the trading needs is controlled by the endowment standard deviation σ_η, which is not easy to pin down. I start with $\sigma_\eta = 1$, so that the difference in exposure to asset risk between the two types of agents, namely $\sigma_\eta(\rho_1 - \rho_2) = 1$, is approximately 3.5 times larger than the per-capita asset risk, which appears to be a fairly large ratio. The implied risk-aversion coefficient is $\gamma = 7.52$. I explain below the sense in which the quantitative results are robust to parameter choice.

I calculate the exact equilibrium price, as well as the linear approximation to the price, for a range of liquidity levels (λ). I also calculate exact and approximated positions.

Rather than reporting the price, I report the more easily interpretable expected excess return on the asset, defined as $m_D/P - r$. Fig. 1 shows that the excess return does, indeed, vary with liquidity, but that even for low levels of liquidity the impact is small. For instance, when $\lambda = 12$, i.e., wait 1 month to trade, the return impact is smaller than 1bp. Here, the excess return decreases with liquidity towards the Walrasian value, but it can also increase, for different parameter configurations.

Portfolio choice, on the other hand, is much more sensitive to liquidity, as Fig. 1 shows. For instance, if one can trade once a month on average ($\lambda = 12$), the lower position is about -1.25, which makes it 14.3% closer to the per-capita supply $\Theta = 1$ than the Walrasian position, -1.63. The same is true, of course, of the high position.[31]

Welfare, like portfolio choice, is also sensitive to liquidity. The bottom left panel of Fig. 1 plots the welfare loss relative to perfect liquidity expressed as a proportion of the total asset market value. For monthly trading, the welfare loss is 24.9% of the asset price.[32] Finally, the annualized turnover, plotted in the last panel, is reasonable at a level close to 100%.

Sensitivity analyses confirm the small price impact of illiquidity. Theorem 1 implies that the price impact can only be significant for sufficiently large γ. I consequently decrease both σ_D and σ_η by a factor of 3 and also increase the Walrasian risk premium to 5%, which results in an absolute risk-aversion coefficient $\gamma = 86.1$—a large value, given the value of the average stock-holding in the economy, 0.57 in this case. The liquidity premium at $\lambda = 12$, in this case, rises more than tenfold to about 10bp, but is still relatively small. At the same time, the welfare impact, a function of $\gamma(\rho_1 - \rho_2)\sigma_D\sigma_\eta$, increases. The turnover, which equals roughly $(\alpha_{12}^{-1} + \alpha_{21}^{-1})^{-1}\frac{\sigma_\eta}{\sigma_D}(\rho_1 - \rho_2)$, changes little. A higher value for $\sigma_\eta(\rho_1 - \rho_2)$—but not σ_D, so as not to push γ down—could generate higher illiquidity return premia, but at the cost of large turnover and relative trade size. (The size of a trade relative to the per-capita supply is very close to the ratio of shock size to per-capita asset risk, $\frac{\sigma_\eta}{\sigma_D}(\rho_1 - \rho_2)$, which, as noted above, is about 3.5 in the base case.)

4. Extensions and discussion

I now turn to discussing more carefully the relationships between this paper and two directly relevant strands of literature: the one studying search in asset markets and the one on transaction costs. I also extend the base-case model of Section 1 to allow for portfolio constraints and transaction costs.[33] In addition to facilitating the discussion, these extensions help clarify a couple of important results in the literature.

4.1. Search literature

The paper is obviously related to the literature on search in asset markets, in that the main friction studied is the same: the impossibility of trading instantaneously. A number of differences with this literature exist, however.

The main difference lies in the fact that this paper allows for a nuanced portfolio choice—all the other papers restrict agents to a binary holding choice.[34] In addition, in the literature trade is bilateral and the price is determined by bargaining. I argue that the different conclusion I draw

[31] The short positions are due to the large hedging needs $\frac{\sigma_\eta}{\sigma_D}(\rho_1 - \rho_2)$. Reducing these hedging needs would further diminish the liquidity impact on the excess return. This feature is helpful, however, for illustrating the implications of prohibiting short sales (Section 4.1).

[32] This welfare loss corresponds to a total transaction cost of about 1.06% of each transaction value in a market with no delays—in other words, to $q = 0.53\%$ in the context of Section 4.2.3.

[33] The set-up is sufficiently flexible and sufficiently tractable to allow for a variety of interesting extensions, which I do not pursue here in order to keep the paper focused. Other possible extensions include allowing for multiple risky assets or an endogenous choice of trading frequency.

[34] In research conducted independently of this paper, Lagos and Rocheteau [24] starts from the premise of quasi-linear preferences and also allows for unrestricted portfolio choice.

concerning the price impact of liquidity has to do with relaxing the portfolio constraints, rather than with the difference in trading mechanism. To do so, in principle I could either introduce risk aversion and portfolio choice in a setting such as DGP or implement portfolio restrictions in the current setting.

The first route requires solving a bilateral-trade model without restrictions on asset holdings, which appears quite difficult. In addition, the outcome would be a range of prices—in fact, an unbounded one, as random matching would result in unbounded asset holdings. One would imagine that the intuition concerning the opposite impact of liquidity on buyers' and sellers' demand curves would continue to hold on average, but the exact outcome is difficult to determine.

In contrast, modifying the current set-up by introducing portfolio constraints and then showing that binding constraints have a significant price impact is relatively easy. In fact, exogenous portfolio constraints, due to agency issues or market features, are quite common, in particular for the type of assets affected by the notion of illiquidity studied here, so such an addition to the model is of independent interest. The most common constraint is the inability to sell short, which is sometimes imposed contractually.[35] I concentrate on this inability for the sake of concreteness.

I give the details of the technical analysis in Appendix A and record the result here.

Proposition 3. *Assume that positions are constrained to satisfy $\theta_k \geqslant 0$ and consider parameters for which the constraint binds. Then the following holds.*

(i) *There exists a value $\underline{\lambda} > 0$ such that the price P increases in λ for $\lambda > \underline{\lambda}$.*
(ii) *If $\mathrm{E}_0[\rho(t) \mid \rho(0) = \rho_k]$ is monotonic in t for all k and in k for all t, then P increases in λ for all λ.*
(iii) *If there are only two types ($J = 2$), then the price is given by*

$$Pr = \kappa_1(\theta_2, \rho_2) - \frac{\alpha_{21}}{r + \lambda + \alpha_{12} + \alpha_{21}} r\gamma(\rho_1 - \rho_2)\sigma_D\sigma_\eta \tag{34}$$

and it increases in λ for all λ.

The main point of the proposition is that, even under the approximation detailed in Theorem 1, the price is sensitive to the liquidity level λ.

The calibrated example of Section 3.4 can also illustrate this point. To that end, I assume that short sales are not allowed and compute the price again for a variety of levels of liquidity, given the parameters in Table 1. As can be seen in Fig. 1, and is reflected in Fig. 2, for low levels of liquidity the optimal holdings are both positive, so the constraints do not bind and the price is approximately equal to the Walrasian price. Beyond a certain threshold, though, the high-correlation agents do not hold any amount of the asset, which fixes the holding of the other agents, too. The excess return decreases as a consequence, to the effect that it becomes about 2% lower than in the illiquid market, as λ goes to infinity. The reason, once again, is the agents' inability to adjust positions to the level of liquidity.[36]

[35] Shorting difficulties have been documented even in the case of listed stocks—see, for instance, Geczy et al. [18] or D'Avolio [9]—which are more liquid than most other securities. They are self evident in the case of private placements.

[36] As an aside, this model can be reinterpreted so that the reasons for trade arise from different beliefs. Then, Proposition 3 adds to the conclusion of Harrison and Kreps [21] and Scheinkman and Xiong [29] that shorting constraints can generate bubbles by showing that the bubble size increases with the future level of liquidity.

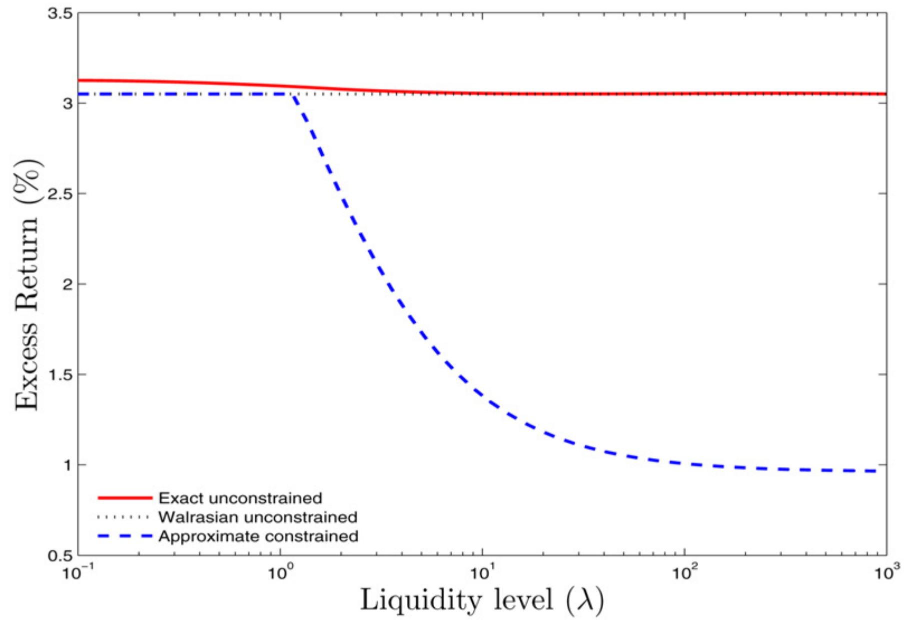

Fig. 2. Excess-return impact of illiquidity in the presence of short-sale constraints (parameters given by Table 1). The continuous line plots the exact unconstrained-holding excess return, the dashed line plots the approximate constrained-holding excess return, while the dotted line shows the Walrasian level.

The numerical example shows that the direct effect of liquidity on the excess return when the positions are constrained can be significant (2%), and at the same time virtually canceled by the effect of the endogenous position adjustment.

Finally, whereas the extension to constraints binding on one side of the market appears to be best motivated when assets are divisible, one can argue that it is not the most direct counterpart to a set-up such as DGP. In DGP no agent, be it buyer or seller, is marginal—in effect, that is a set-up with indivisible assets. However, qualitatively identical results obtain when, in the DGP set-up, agents gain access to a centralized market with intensity λ. Provided that $\alpha_{12} > \alpha_{21}$,[37] the prevailing price is given by

$$Pr = \frac{\kappa(\theta_2, \rho_2) - \kappa(\theta_1, \rho_2)}{\theta_2 - \theta_1} - \frac{\alpha_{21}}{r + \lambda + \alpha_{12} + \alpha_{21}} r\gamma(\rho_1 - \rho_2)\sigma_D\sigma_\eta. \tag{35}$$

This result shows clearly that the small liquidity impact on price is due not to the centralized-market feature in the current paper, but to the free choice of holding size in a divisible asset.

4.2. Transaction costs

The intuition behind the reduced price effect of liquidity in the sense of infrequent trading opportunities is useful for other notions of illiquidity—in particular, for exogenous transaction costs, such as brokerage fees. In this context, too, the (partial) cancellation of the illiquidity effect on the demands of buyers and sellers results in a muted price impact, as shown below. I start by comparing the two types of illiquidity, after which I give a brief review of the main results in the transaction-cost literature. Section 4.2.3 presents a modeling extension that incorporates

[37] This means there are more high-valuation than low-valuation agents in the economy. As a consequence, the price is set by the buyers, just as it is, largely, in DGP, because of the condition $s < \lambda_u/(\lambda_u + \lambda_d)$ in that paper's notation.

transaction costs in the current framework and helps clarify some of these results. I discuss the relationship with the literature in Section 4.2.4.

4.2.1. Transaction costs and trading delays

Clear fundamental differences exist between a brokerage fee, on one hand, and the legal prohibition to trade for a while, or the time-consuming phone search for a trading counterparty together with the due-diligence process, on the other. An immediate consequence of the different natures of these frictions can be stated as follows: Delays are costly only because they prevent agents from trading in order to achieve the optimal portfolio, not because they make trade costly. The size of the costs induced by waiting is endogenous, as it depends on the market conditions, endowments, preferences, and so forth. Transaction costs, on the other hand, impose exogenous losses whenever a trade is made. (In addition, of course, they also prevent the agent from holding optimal positions and may affect the timing of her trades, similar to delays.)

The distinctions between trading delays and transaction costs have implications for equilibrium prices. From this point of view, perhaps economies outside steady state provide the most obvious example: Search naturally delays the adjustment following a shock,[38] but transaction costs, which allow for instantaneous trade, are much less likely to do so.

Another difference in price implications is related directly to the fact that only transaction costs impose losses when trading. For this reason, the equilibrium price reflects the current level of transaction costs, as in Eq. (42) in Section 4.2.3, even if it does not reflect the search efficiency (Eq. (25)). A similar conclusion with respect to the current illiquidity level holds concerning optimal portfolio choice.

Furthermore, in the case of transaction costs, the timing of trades is endogenous: the agent chooses to incur the costs if and only if the utility gain from adjusting her portfolio is sufficiently large. This feature has two different implications when compared with search frictions. First, the deviation from the Walrasian portfolio choice is bounded by the size of the transaction cost, not by the severity of the need for trading: severely disadvantageous holdings can always be avoided. (This implication is captured by the fact that σ_η and ρ do not influence the last term in Eqs. (43)–(44).) Second, a stronger link exists between the timing of a trade and the gain from trade. With fixed transaction costs, this link also implies a tighter relationship between the timing of trades and the more easily observable trade sizes (see Footnote 41).

4.2.2. A brief literature review

An extensive body of work is dedicated to the study of transaction costs in asset markets. For lack of space, I limit my discussion to some of the most important of these papers. In Section 4.2.4, I interpret their results in the context of my own model.

Amihud and Mendelson [1] proposes a general-equilibrium model in which risk-neutral agents must exit the market and sell to newly arriving agents, and finds that the required excess return on an asset equals the product of the asset's turnover and the proportional transaction cost. Constantinides [8], on the other hand, shows in a partial-equilibrium setting that an agent's choice of trading frequency results in a price impact of transaction costs that is an order of magnitude smaller than Amihud and Mendelson [1] suggests. Furthermore, Constantinides [8] assigns the result to the small welfare loss due to transaction costs. Vayanos [31] revisits the issue by

[38] DGP work out a detailed example. Weill [35] provides another good example in a setting with a market maker, who "leans against the wind" in anticipation of a shock. The point can also be made in the current framework, but I do not pursue this line.

building a general-equilibrium, overlapping-generations model with long-lived agents who trade optimally throughout their deterministic lifetimes. He shows how transaction costs affect agents' trading decisions, and therefore risk exposures, and that this effect can largely cancel the direct effect of a lower trading frequency, thus yielding a low price impact. Huang [22] is able to generate somewhat larger price impacts in an OLG model in which lifetime is uncertain. Finally, unlike the other papers, Lo et al. [25] studies a general-equilibrium setting with fixed transaction costs and high-frequency transaction needs, and exhibits parameter configurations resulting in price impacts of the same order as the transaction costs.

4.2.3. A model extension

I now incorporate transaction costs into the basic model of Section 1. In order to keep the analysis as simple as possible, I concentrate on the case in which transaction costs are the only friction.[39] Thus, all agents can trade instantly when they wish to do so, but they have to pay transaction costs proportional to the number of shares traded. Specifically, the buyer pays $P + q$ and the seller receives $P - q$ per share, for some $q \geqslant 0$.

It follows that, for any agent buying,

$$P + q = \mathrm{E}_0\left[\int_0^\infty e^{-rs}\kappa_1\big(\theta(s), \rho(s)\big)\,ds \,\middle|\, \theta(0) = \theta_k^b,\ \rho(0) = \rho_k\right], \tag{36}$$

while, for any seller,

$$P - q = \mathrm{E}_0\left[\int_0^\infty e^{-rs}\kappa_1\big(\theta(s), \rho(s)\big)\,ds \,\middle|\, \theta(0) = \theta_k^s,\ \rho(0) = \rho_k\right]. \tag{37}$$

Note that for any type ρ_i there are two positions an agent would trade to, θ_i^b if he bought and θ_i^s if he sold. Furthermore, an agent of type j holding the optimal position of an agent of type $i \neq j$ may not trade when given the opportunity, if the transaction costs are large. If the transaction cost q is small enough, however, such an agent always trades if he can. Consequently, in this case, one of Eqs. (36), (37) holds for any agent in the market at time t.

In order to write the demand schedules, let τ denote the arrival time of the first type change after time 0 for the agent considered. Note that a buyer currently considering acquiring a marginal unit either will save $P + q$ if he turns out to want to buy at τ—because his type decreased—or will receive $P - q$ if he sells next time. Consequently,

$$P + q = \mathrm{E}_0\left[\int_0^\tau e^{-rs}\kappa_1\big(\theta_k^b, \rho_k\big)\,ds\right] + \mathrm{E}_0\big[e^{-r\tau}\big(P + q1_{(\text{buy at }\tau)} - q1_{(\text{sell at }\tau)}\big)\big],$$

or, using the exponential arrival times and rewriting,

$$P = \frac{1}{r}\kappa_1\big(\theta_k^b, \rho_k\big) - q - 2q\frac{\alpha^{k+}}{r}. \tag{38}$$

Here $\alpha^{k+} = \sum_{j<k}\alpha_{kj}$, i.e., α^{k+} is the intensity with which type ρ_k switches to a type $\rho > \rho_k$, which induces the agent to want to sell. The analogous equation for a seller is

[39] Appendix A contains the results for the general case.

$$P = \frac{1}{r}\kappa_1\left(\theta_k^s, \rho_k\right) + q + 2q\frac{\alpha^{k-}}{r}, \tag{39}$$

with $\alpha^{k-} = \sum_{j>k} \alpha_{kj}$, and the resulting demand equations are

$$\theta_k^b = \frac{1}{\gamma\sigma_D^2}\left(\frac{m_D}{r} - \left(P + q + 2q\frac{\alpha^{k+}}{r}\right)\right) - \frac{\sigma_\eta}{\sigma_D}\rho_k, \tag{40}$$

$$\theta_k^s = \frac{1}{\gamma\sigma_D^2}\left(\frac{m_D}{r} - \left(P - q - 2q\frac{\alpha^{k-}}{r}\right)\right) - \frac{\sigma_\eta}{\sigma_D}\rho_k. \tag{41}$$

These equations are similar to (23). For the buyer, the two differences are that (i) he pays more for the asset—$P + q$ instead of P—and (ii) if he chooses to sell next time in the market he will incur round-trip transaction costs. For both reasons, he demands a lower price today. Analogous reasoning yields the higher price demanded by the seller. Note that higher transaction costs increase the demand of the sellers and decrease the demand of the buyers.

Aggregating (40) and (41) yields the equilibrium price. To write it, let μ_k^b denote the total mass of agents of type ρ_k who hold θ_k^b and define μ_k^s analogously. Also, let $\mu^b = \sum_k \mu_k^b$ and $\mu^s = 1 - \mu^b$. Thus, μ^b (μ^s) is the total mass of all agents who bought (sold) last time they traded. With θ_k^W denoting the position chosen in the Walrasian market, the following holds.

Proposition 4. *There exists $\bar{q} > 0$ such that, for transaction costs $q \leqslant \bar{q}$, in steady state*

$$P = P^W - q\left(\mu^b - \mu^s\right) \tag{42}$$

and θ_k^b and θ_k^s, given by (40) *and* (41)*, can be written as*

$$\theta_k^b = \theta_k^W - \frac{2q}{\gamma\sigma_D^2}\left(\mu^s + \frac{\alpha^{k+}}{r}\right), \tag{43}$$

$$\theta_k^s = \theta_k^W + \frac{2q}{\gamma\sigma_D^2}\left(\mu^b + \frac{\alpha^{k-}}{r}\right). \tag{44}$$

4.2.4. Discussion

Transaction costs impact both the equilibrium portfolios and the equilibrium price. In particular, as is natural, the buyers' asset holdings decrease with transaction costs, whereas the sellers' increase. As a consequence, transaction costs reduce trade volume.

The price also depends on transaction costs. Importantly, however, the price impact of transaction costs equals the difference between the discount buyers require and the premium sellers require. This (partial) offsetting of opposite transaction-costs effects on demand schedules reduces the magnitude of the effect on the price. Viewed through the prism of the effect on demand schedules, the sign of the price effect is also natural: the price decreases with transaction costs if and only if more buyers than sellers have to be attracted to the market ($\mu^b > \mu^s$). While a buyer's and a seller's demands are impacted symmetrically, a larger number of buyers would make their aggregate demand more sensitive to transaction costs.

Interestingly, the entire price effect is due to the current transaction cost, and is therefore small relative to the price, as future costs are not relevant. This conclusion is in contrast with earlier results, e.g., Amihud and Mendelson [1].[40] The intuition for my result is in line with the notion that all demand schedules are affected, and not in the same way. In particular, future costs are important because an agent may complete a round-trip transaction, and so would the agent with whom she would trade, and so forth. This effect reduces the demands of the current buyers while increasing those of the current sellers. In a stationary environment, the number of buyers who turn sellers at any point in time is equal to the number of sellers becoming buyers, which results in equal and opposite, therefore precisely offsetting, adjustments to the aggregate buyer and seller demands. (Note that the same is true when illiquidity takes the form of trading delays, the only difference being that delays do not impose any cost when trading, unlike transaction costs. The price is therefore not affected at all.)

The lack of price impact of future, amortized transaction costs does not obtain in the literature because of the life cycle of the agents: In the current paper, agents are not forced to sell at some point regardless of the price. A reasonable conjecture is that if some trading was due to agents' life cycles, then the frequency of such trading would depress the price. In Appendix A I confirm this conjecture by showing that if agents enter and exit the economy at rate π, then, under appropriate conditions,

$$P = P^W - q\left(\mu^b - \mu^s\right) - 2q\frac{\pi}{r}. \tag{45}$$

Eq. (45) refines the results of Amihud and Mendelson [1]: The price is diminished by the sum of amortized future transaction costs, but only to the extent that the costs are incurred when the trader is not marginal. Thus, even if the agents may trade frequently, the price impact $P^W - P$ of transaction costs is small if π is small, that is, if investors are in the market for the long run. Distinguishing between the total frequency with which the marginal share is traded and the frequency with which it is forced to be traded is thus important for the purpose of determining the required return. A clear empirical implication is that, for assets with similar turnovers, transaction costs should have less price impact on the one in which agents have longer market horizons—equivalently, trade more for portfolio rebalancing reasons. This conclusion is consistent with the low impacts found by Constantinides [8], where rebalancing is the only reason for trade, and Vayanos [31], where the market horizon is very long, as well as the moderate impact of Huang [22], where the liquidity shocks are quite frequent ($\pi = 0.5$).

Of the transaction-cost literature, Lo et al. [25], henceforth LMW, is arguably the most closely related paper, as it studies general equilibrium with infinitely lived agents that trade repeatedly in a CARA-normal world to hedge endowment risk. The two noteworthy differences between the two set-ups are that LMW considers fixed transaction costs, rather than trading delays or proportional transaction costs, and that the hedging needs in LMW (given by X, in the paper's notation) follow a (continuous) random-walk process, rather than a mean-reverting (and discrete-valued) process.

In the current model, the addition of fixed transaction costs makes no difference to demand curves, and therefore price, as long as the costs are small enough not to affect the timing of trades. Further, in the context of LMW, the analogous approximation to the one I use here yields

[40] In this connection, see also Vayanos [31], where there is also only a small price impact, and not necessarily of negative sign. The current framework shows clearly that the small transaction-cost impact is not due to the adjustment in trading frequency of the marginal share, as the results of Constantinides [8] have been interpreted to show. Rather, it is due to the adjustment in investors' risk exposures, as suggested by Vayanos [31].

demand curves, and therefore prices, unaffected by future transaction costs.[41] (One can prove this assertion by considering a deviation consisting of buying and holding forever an additional ϵ number of shares, for small $\epsilon > 0$, and using the fact that the risk exposure is a martingale in LMW.) When moving away from this approximation, however, the fact that the marginal utility is not linear comes into play: agents suffer more from adding risk to their original exposure (as the per-capita risk exposure is positive) than they benefit from reducing risk. This imbalance decreases all agents' demands, and consequently the price. A similar effect obtains in the current paper, but the lack of perfect symmetry (for instance, in the case $J = 2$, $\alpha_{12} = \alpha_{21}$ is not required) can induce an illiquidity premium as well as discount.

Finally in this connection, LMW is able to display reasonable parameters yielding non-negligible illiquidity premia. Naturally, these premia increase with the size σ_X of the hedging motive. Despite using the same common parameters, I find smaller price impacts for what I argue to be empirically relevant parameter values. However, at the cost of excessively large trade sizes and turnover, sufficiently large hedging needs $(\rho_1 - \rho_2)\sigma_\eta$ can generate significant premia in this paper, too.

The last observation in the context of transaction costs concerns welfare. As I note in Sections 3.3 and 3.4, autarky can lead to a significant welfare loss relative to the Walrasian market. Consequently, transaction costs can also have an important welfare impact, despite having a minor effect on the price.

5. Conclusion

This paper studies portfolio choice and pricing in markets in which trading may take place with considerable delay. Examples include private placements, which may have to be held without trading for one or more years, as well as assets for which finding an appropriate buyer or seller may require lengthy search, such as small corporate and municipal bond issues and shares in firms recently emerged from Chapter 11 proceedings.

I derive closed-form price and asset-holding expressions, based on an approximation supported by numerical results. I find that the liquidity level has a strong impact on portfolio choice. For instance, when expecting future trading to be difficult, an institution with current high value for a particular corporate bond—say, due to low correlation with the rest of its portfolio—should buy a smaller amount of the bond than it would in a perfect market. The reason is that the institution may have to continue maintaining its position for a while after its value for the bond diminishes. Similarly, if its value from the asset could increase in the future, the institution should hold a larger amount when the market is illiquid. A clear empirical implication is that smaller blocks are traded in illiquid markets.

Second, the paper highlights a simple general intuition for a mitigated price impact of illiquidity when portfolio choice is unrestricted. This intuition is that agents with relatively high asset valuation diminish their demands, whereas those with relatively low valuation increase theirs, as liquidity worsens. These opposite demand adjustments result in a smaller effect on the aggregate demand than if one type of demands were fixed. This effect obtains both in the context of infre-

[41] Current (fixed) transaction costs affect only the decision of whether to trade today. Higher such costs make agents postpone trades, in general. With random-walk needs for trade, this means that higher transaction costs are deterministically linked to larger trade sizes. In the context of search, higher illiquidity translates automatically into less frequent trades, but, even with random-walk needs for trade, higher illiquidity only implies larger trades on average. Trade-size variance actually increases with the illiquidity level.

quent trading and in the context of transaction costs. Binding portfolio constraints, on the other hand, prevent some agents' demands from adjusting. Binding shorting prohibitions, for instance, can generate a significant illiquidity price discount. This observation refines earlier results in the search-in-financial-markets literature, showing that, while such a discount can obtain, it is likely to require asset indivisibility or other restrictions on holdings. In addition, such a discount could also obtain if sellers are over-represented in the market, perhaps due to costly participation or other institutional reasons, or if marginal utilities are sufficiently non-linear.

In the context of transaction costs, the reasoning here shows that, contrary to the conclusion of the literature, the frequency of trade per se does not determine the price discount. Rather, the relevant frequency is that with which a trader is forced to liquidate her position, and therefore trade without being marginal. Consequently, turnover is unlikely to be the correct empirical quantity to use for calculating discounts due to transaction costs. Finally, the findings of the paper belie the inference made in the literature (e.g., Constantinides [8]) that the small price impact is due to the insensitivity of welfare to liquidity.

Acknowledgment

I am grateful for discussions with Domenico Cuoco, Darrell Duffie, Jennifer Huang, Christine Parlour, Lasse Pedersen, Rohit Rahi, Guillaume Rocheteau, Dimitri Vayanos, Pierre-Olivier Weill, and Jean-Pierre Zigrand, as well as for comments from seminar participants at Wharton, the ASAP conference at the LSE, the 2006 WFA Meetings at Keystone, the SITE Conference on Imperfect Markets, the SIFR Conference on Institutions and Liquidity, the 2007 Frontiers of Finance Conference, University of Lausanne, the 2007 European Winter Finance Conference, the 2007 SAET Conference, and the 2007 LAEF Trading Frictions in Asset Markets Conference. I am solely responsible for all errors.

Appendix A. General trading times

The fact that trading times follow independent Poisson processes guarantees stationarity and provides simple closed-form solutions. The model, however, can easily incorporate more complex distributions of trading dates for each agent, as well as time variation in other parameters, such as α_{ij}. The key is that the formulation (18) of the objective of an agent continues to hold.

Let $\lambda_t \in [0, \infty)$ be measurable with respect to the Lebesgue measure on the real line. In addition, let there be a (possibly infinite) sequence of stopping times $T_1^o \leqslant T_1^c < T_2^o \leqslant T_2^c < \cdots$ such that all agents (or a random subsample) have continuous access to a centralized market between dates T_i^o and T_i^c.

It is not required that all agents' trading times have the same distribution, or that all agents' types have the same distribution, or that the type-transitions α_{ij} be time-independent. It suffices that, for any times $s \geqslant t$,

$$\mathrm{E}\big[\rho(s) \,\big|\, \text{market access at } t\big] = \bar{\rho}, \tag{A.1}$$

$$\mathrm{E}\big[\theta(s) \,\big|\, \text{market access at } t\big] = \Theta. \tag{A.2}$$

Condition (A.2), which is endogenous, can be ensured by making the primitive assumption that the identity of agents allowed to trade at any time is independent of their type and history. (This assumption can be further weakened at the expense of additional complexity.)

It follows immediately that aggregating Eq. (19) yields the same price as when all agents trade continuously, i.e., the Walrasian price. Of course, the quantities $\tilde{\mu}$ are different from the case in the body of the paper, and so are the optimal positions θ_k, but the intuition for the determination of the position choice is very similar to that in the basic model: knowing that the positions may not be adjustable for a certain period of time, agents incorporate in their decision their future benefits from ownership. Thus, an agent with high present need for the asset takes a smaller position than in a perfect market, given that she may be forced to keep the same position even if her need diminished.

Proof of Theorem 1. Suppose first that a solution to Eqs. (11), (12), and (14) exists. Let $V(W,\theta,\rho)$ be the value function. Given the liquid-wealth dynamics (5), V must obey the following HJB equation:

$$\begin{aligned}0=\sup_{\bar c\in\mathbb{R}}\Bigg\{&-e^{-\gamma\bar c}+V_W(W,\theta,\rho_j)(rW-\bar c+\theta m_D+m_\eta)\\&+\frac{1}{2}V_{WW}(W,\theta,\rho_j)\big(\theta^2\sigma_D^2+\sigma_\eta^2+2\rho_j\theta\sigma_D\sigma_\eta\big)-\beta V(W,\theta,\rho_j)\\&+\sum_{l\neq j}\alpha_{jl}\big[V(W,\theta,\rho_l)-V(W,\theta,\rho_j)\big]\\&+\lambda\sup_{\bar\theta}\big[V\big(W-P(\bar\theta-\theta),\bar\theta,\rho_j\big)-V(W,\theta,\rho_j)\big]\Bigg\}.\end{aligned}\tag{A.3}$$

Letting

$$\bar a=\frac{1}{r}\left(\frac{\log r}{\gamma}+m_\eta-\frac{1}{2}r\gamma\sigma_\eta^2-\frac{r-\beta}{r\gamma}\right),\tag{A.4}$$

conjecture

$$V(W,\theta,\rho)=-e^{-r\gamma(W+a(\theta,\rho)+\bar a)}.\tag{A.5}$$

Since

$$\begin{aligned}&V_W(W,\theta,\rho)=-(r\gamma)V(W,\theta,\rho),\\&V_{WW}(W,\theta,\rho)=(r\gamma)^2V(W,\theta,\rho),\end{aligned}$$

the first-order condition with respect to consumption implies an optimal consumption rate of

$$\bar c=-\frac{\log(r)}{\gamma}+r\big(W+a(\theta,\rho)+\bar a\big).\tag{A.6}$$

Using this value in (A.3) yields (11). It remains to show that the transversality condition is satisfied by the proposed strategy. Given that θ is bounded, it suffices to see that

$$e^{-r\gamma(W_t+a(\theta(t),\rho(t))+\bar a)}=e^{-(\beta-r)(T-t)}\mathrm{E}_t\big[e^{-r\gamma(W_T+a(\theta(T),\rho(T))+\bar a)}\big],$$

which can be verified directly exactly as in DGP.

The approximations follow immediately from the fact that the equilibrium value-function coefficients $a_j(\theta_k)$ and price are bounded as $\gamma\to 0$ keeping $\gamma\sigma_D\sigma_\eta$ fixed.

As for the Walrasian equilibrium, it is characterized by

$$-ra_j\big(\theta_j^W\big)=\sum_l\alpha_{jl}\sup_{\bar\theta}\frac{e^{-r\gamma(-P^W(\bar\theta-\theta_j^W)+a_l(\bar\theta)-a_j(\theta_j^W))}-1}{r\gamma}-\kappa\big(\theta_j^W,\rho_j\big)$$

$$= \sum_l \alpha_{jl} \frac{e^{-r\gamma(-P^W(\theta_l^W - \theta) + a_l(\theta_l^W) - a_j(\theta_j^W))} - 1}{r\gamma} - \kappa(\theta_j^W, \rho_j), \tag{A.7}$$

together with (14):

$$\begin{aligned} P^W &= a_j'(\theta_j^W) \\ &= \frac{1}{r}\kappa_1(\theta_j^W, \rho_j), \end{aligned} \tag{A.8}$$

where the second equality follows by differentiating (A.7). Now the market-clearing condition yields the price P^W, and then (A.8) the optimal quantities θ_j^W.

Note also that the bound M on $|\theta|$ can be chosen as any number larger than $\max_i |\theta_i^W|$.

Finally, I give a sketch of the argument that the system (11), (12), and (14) admits a real solution. The argument runs on the more primitive agent optimization problem and uses standard fixed-point tools for dynamic programming. The main steps are as follows. (As ρ takes values in a finite set, all the properties of functions Y below that hold uniformly in θ on some set A for any value of ρ hold uniformly in θ. From now on, I refer to Y as if it were only a function of θ.)

1. Fixing a price P, define an operator O by

$$\begin{aligned} &(OY)(\theta, \rho) \\ &\quad = \inf_{\bar{c}, \bar{\theta}} \left\{ \mathrm{E}\left[\int_0^\tau e^{-\gamma \bar{c}_t - \beta t}\, dt + e^{-\beta\tau - r\gamma(W_\tau + P(\theta - \bar{\theta}))} Y(\bar{\theta}, \rho(\tau)) \,\middle|\, \rho(0) = \rho, \right.\right. \\ &\quad\quad \left.\left. W_0 = 0 \right] \right\}. \end{aligned}$$

   I need to show that it admits a fixed point.
2. O is monotone: $Y^1 > Y^2 \Rightarrow OY^1 > OY^2$.
3. OY is convex in the first argument for all Y.
4. There exists $Y > 0$ such that $OY \leqslant Y$. For instance, pick $Y^0(\theta, \rho) = e^{P\theta + A\theta^2 + B}$ for A and B large enough. For all $n > 0$, let $Y^n = O^n Y^0$.
5. There exist $u > 0$ and v such that $Y^n(\theta, \rho) > e^{u\theta^2 + v}$.
6. Y^n is a decreasing sequence of finite-valued positive convex functions, bounded below by a continuous function. Let $\bar{Y}(\theta, \rho) = \lim_n Y^n(\theta, \rho)$—it is a (weakly) convex function on the real line, therefore continuous.
7. Given the monotonicity in n of the finite-valued convex family $\{Y^n\}$, its lower bound, and its finite upper bound, $(Y^n)'$ is bounded uniformly on compacts. Consequently, $\{Y^n\}$ is equicontinuous on compacts. This ensures that the convergence of Y^n to $\bar{Y}$ is uniform, and therefore $\bar{Y}$ is a fixed point of O.
8. As $P \to \infty$, the optimal $\theta \to -\infty$, and vice versa. Consequently, P exists that clears the market. □

Proof of Proposition 2. For part (i), note from Eqs. (26) and (27) that

$$\theta_k = \theta_k^W - \frac{\sigma_\eta}{\sigma_D}\left(\sum_j \tilde{\mu}_{jk}\rho_j - \rho_k\right). \tag{A.9}$$

Since ρ_1 and ρ_J are the maximum, respectively minimum, values that ρ can take, $\theta_1^W < \theta_k < \theta_J^W$.

Furthermore, the quantities $\tilde{\mu}_{jk}$ are rational functions of λ, and consequently so are θ_k. Therefore, the quantities θ_k have only a finite number of local maxima or minima. Consider $\underline{\lambda}$ higher than all such local extrema.

Up to terms in $O(\lambda^{-2})$ for large λ, it is easily seen that, with $j \neq k$,

$$\tilde{\mu}_{kk} \simeq 1 - \frac{\sum_{i \neq k} \alpha_{ki}}{\lambda}, \tag{A.10}$$

$$\tilde{\mu}_{jk} \simeq \frac{\alpha_{kj}}{\lambda}, \tag{A.11}$$

and, consequently,

$$\theta_k \simeq \theta_k^W - \frac{1}{\lambda} \frac{\sigma_\eta}{\sigma_D} \sum_{j \neq k} \alpha_{kj} (\rho_j - \rho_k). \tag{A.12}$$

Since θ_k is monotonic in λ, the sign of its dependence is given by that of $\sum_{j \neq k} \alpha_{kj}(\rho_j - \rho_k)$. It is clear that θ_1 decreases, while θ_J increases with λ for $\lambda > \underline{\lambda}$.

For part (ii), write

$$\theta_k = \Theta - \frac{\sigma_\eta}{\sigma_D} \int_0^\infty f(\lambda, t) \big(\mathrm{E}_0 \big[\rho(t) \bigm| \rho(0) = \rho_k \big] - \bar{\rho} \big) \, dt, \tag{A.13}$$

where

$$f(\lambda, t) = (r + \lambda) e^{-(r+\lambda)t}.$$

Note that $\mathrm{E}_0[\rho(t) \mid \rho(0) = \rho_k] - \bar{\rho}$ tends to 0 monotonically as $t \to \infty$, and that the distribution on $[0, \infty)$ with pdf $f(\lambda_1, t)$ first-order stochastically dominates the one with pdf $f(\lambda_2, t)$ whenever $\lambda_1 < \lambda_2$.

It follows immediately that θ_k is monotonic in λ. In particular, θ_k increases with λ if and only if $\mathrm{E}_0[\rho(t) \mid \rho(0) = \rho_k] - \bar{\rho}$ increases, which is equivalent to $\mathrm{E}_0[\rho(t) \mid \rho(0) = \rho_k] \leqslant \bar{\rho}$; thus, if $\theta_k \geqslant \Theta$.

Finally, for part (iii), with $\bar{k} = 3 - k$ one calculates explicitly the quantities

$$\tilde{\mu}_{kk} = \frac{r + \lambda + \alpha_{\bar{k}k}}{r + \lambda + \alpha_{12} + \alpha_{21}}, \tag{A.14}$$

$$\mu_{k\cdot} = \frac{\alpha_{\bar{k}k}}{\alpha_{12} + \alpha_{21}}, \tag{A.15}$$

$$\mu_{\bar{k}k} = \frac{\alpha_{12}\alpha_{21}}{(\alpha_{12} + \alpha_{21})(\lambda + \alpha_{12} + \alpha_{21})}. \qquad \square \tag{A.16}$$

The following result analyzes the different effects of past and future liquidity levels.

Proposition 5. *Consider a market with two types ($J = 2$), and fix a certain distribution of holdings at time 0, with the types distributed as in steady state. Let the meeting intensity be given by $\lambda(t) = \lambda$ for $t < T$ and $\lambda(t) = \lambda'$ for $t \geqslant T$. Then*

(i) *$\theta_1(t)$ decreases and $\theta_2(t)$ increases with λ' for all $t \leqslant T$;*
(ii) *$\mu_{12}(t) + \mu_{21}(t)$ decreases with λ for $t > T$, provided that $T > \underline{T}$ for some $\underline{T}$ independent of λ;*

(iii) $\mu_{12}(t)+\mu_{21}(t)$ *decreases with* λ *for all* t, *provided that* $\mu_{12}(0)+\mu_{21}(0)\geqslant\mu^*_{12}(\lambda)+\mu^*_{21}(\lambda)$, *where* $\mu^*_{jk}(\lambda)$ *is the steady-state value of* μ_{jk} *corresponding to* λ.

The first part of the proposition shows that improved future liquidity is anticipated by more extreme positions, and therefore larger trades, today. The second and third parts show that the number of agents trading in the future is lower if current liquidity is high, capturing the intuition that current trading reduces the number of agents needing to trade in the future. The only case in which this intuition might fail is when the economy starts with a low number of traders, which increases over time towards steady state. An increase in liquidity, then, can result in more traders in the beginning, though not if enough time elapses. The conditions provided rule this case out.

Proof of Proposition 5. For part (i), start from

$$\mathrm{E}_t\left[\int_t^\tau e^{-r(s-t)}\kappa_1\big(\theta_k(t),\rho(s)\big)\,ds\;\Big|\;\rho(t)=\rho_k\right]=\mathrm{E}_t\left[\int_t^\tau e^{-r(s-t)}\kappa_1(\Theta,\bar\rho)\,ds\right].$$

Using the linearity of κ_1, it follows that

$$\begin{aligned}\theta_k(t)&=\Theta+\left(\bar\rho_k-\frac{\mathrm{E}_t[\int_t^\tau e^{-r(s-t)}\rho(s)\mid\rho(t)=\rho_k]}{\mathrm{E}_t[\int_t^\tau e^{-r(s-t)}]}\right)\frac{\sigma_\eta}{\sigma_D}\\&=\Theta+\left(\bar\rho_k-\int_t^\infty\Psi(s;\lambda,\lambda')\mathrm{E}_t\big[\rho(s)\;\big|\;\rho(t)=\rho_k\big]\,ds\right)\frac{\sigma_\eta}{\sigma_D},\end{aligned}$$

where Ψ is defined by the last equation, and is the only quantity that depends on λ'. To finish the proof, make use again of the first-order stochastic dominance w.r.t. λ', for fixed λ, and the monotonicity of $\mathrm{E}_t\,[\rho(s)\mid\rho(t)=\rho_k]$ in s.

For part (ii), let $\bar\mu=\mu_{12}+\mu_{21}$. For $t\leqslant T$,

$$\bar\mu_t=\bar\mu^*(\lambda)+e^{-(\lambda+\alpha_{12}+\alpha_{21})t}\big(\bar\mu_0-\bar\mu^*(\lambda)\big),\tag{A.17}$$

while for $t\geqslant T$,

$$\bar\mu_t=\bar\mu^*(\lambda')+e^{-(\lambda'+\alpha_{12}+\alpha_{21})(t-T)}\big(\bar\mu_T-\bar\mu^*(\lambda')\big),\tag{A.18}$$

where $\bar\mu^*(\lambda)$ is the steady-state value of $\bar\mu$ corresponding to λ. For $t\geqslant T$, the dependence on λ of $\bar\mu_t$ is the same as that of

$$\bar\mu_T=\bar\mu^*(\lambda)+e^{-(\lambda+\alpha_{12}+\alpha_{21})T}\big(\bar\mu_0-\bar\mu^*(\lambda)\big).$$

Since

$$\frac{\partial}{\partial\lambda}\bar\mu_T=\frac{\partial}{\partial\lambda}\bar\mu^*(\lambda)\big(1-e^{-(\lambda+\alpha_{12}+\alpha_{21})T}\big)-Te^{-(\lambda+\alpha_{12}+\alpha_{21})T}\big(\bar\mu_0-\bar\mu^*(\lambda)\big),\tag{A.19}$$

part (ii) follows from the fact that $\bar\mu^*$ changes more slowly than the exponential function, for T large enough.

As for part (iii), it also follows from (A.19) because $\frac{\partial}{\partial\lambda}\bar\mu^*(\lambda)\leqslant 0$ and $1-e^{-(\lambda+\alpha_{12}+\alpha_{21})T}\geqslant 0$. □

Proof of Proposition 3. Suppose that, at all times, every position θ_k must satisfy $\theta_k \geqslant \underline{\theta}$.[42] Also, let $\hat{\mu}_{jk} = \mu_k \tilde{\mu}_{jk}$. For any position $\theta_k > \underline{\theta}$ chosen optimally, the pricing equation (22) holds:

$$\mu_k Pr = \sum_j \hat{\mu}_{jk} \kappa_1(\theta_k, \rho_j). \tag{A.20}$$

Fixing λ, let us assume that $\theta_k > \underline{\theta}$ if and only if $k > k_0$—this condition, which need not be true in general, holds under the assumptions of the proposition, as shown below. Aggregating (A.20) over the values of k for which it holds, i.e., $k > k_0$, yields

$$\left(\sum_{k>k_0} \mu_k\right) Pr = \sum_{j,k>k_0} \hat{\mu}_{jk} \kappa_1(\theta_k, \rho_j),$$

or

$$Pr = \kappa_1\left(\left(\sum_{k>k_0} \mu_k\right)^{-1}\left(\sum_{k>k_0} \mu_k \theta_k\right), \left(\sum_{k>k_0} \mu_k\right)^{-1}\left(\sum_{j,k>k_0} \hat{\mu}_{jk} \rho_j\right)\right).$$

While lengthy, the expression above is as natural as the one obtaining without constraints: $Pr = \kappa_1(\Theta, \bar{\rho})$. The price is given by the per-capita supply of assets held away from the constraints, and the average discounted type among unconstrained agents.

It is clear that the price dependence on the liquidity level is determined by the term $\sum_{j,k>k_0} \hat{\mu}_{jk} \rho_j$, on which I now concentrate.

For part (i), note first that, for λ high enough, $|\theta_k - \theta_k^W|$ is sufficiently small for all k that $\theta_1 \leqslant \theta_2 \leqslant \cdots \leqslant \theta_J$, so that k_0 exists as assumed. Up to terms in $o(\lambda^{-1})$,

$$\begin{aligned}
\sum_{j,k>k_0} \hat{\mu}_{jk} \rho_j &= \sum_{j,k>k_0} \mu_k \tilde{\mu}_{jk} \rho_j \\
&\simeq \sum_{k>k_0} \mu_k \left(\sum_{j\neq k} \frac{\alpha_{kj}}{\lambda}(\rho_j - \rho_k) + \rho_k\right) \\
&= \sum_{k>k_0} \mu_k \rho_k + \frac{1}{\lambda} \sum_{k>k_0, j\neq k} \mu_k \alpha_{kj} (\rho_j - \rho_k).
\end{aligned}$$

The ranking of ρ_j and $\sum_{k,j\neq k} \mu_k \alpha_{kj}(\rho_j - \rho_k) = 0$ imply that $\sum_{k>k_0, j\neq k} \mu_k \alpha_{kj}(\rho_j - \rho_k) > 0$.

For part (ii), note that the assumed monotonicity in k implies monotonicity in k of θ_k, ensuring the existence of k_0. Write

$$\begin{aligned}
\sum_{j,k>k_0} \hat{\mu}_{jk} \rho_j &= \sum_{j,k>k_0} \mu_k \tilde{\mu}_{jk} \rho_j \\
&= (r+\lambda) \sum_{k>k_0} \mu_k \int_0^\infty e^{-(r+\lambda)t} \mathrm{E}_0\big[\rho(t) \bigm| \rho(0) = \rho_k\big]\, dt.
\end{aligned}$$

Since $\mathrm{E}_0[\rho(t) \mid \rho(0) = \rho_k]$ decreases if and only if $\rho_k \geqslant \bar{\rho}$, and $\rho_{k_0} > \bar{\rho}$ by virtue of the fact that $\underline{\theta} < \Theta$, $\sum_{k\leqslant k_0} \mu_k \mathrm{E}_0[\rho(t) \mid \rho(0) = \rho_k]$ decreases with t. Given that $\sum_k \mu_k \mathrm{E}_0[\rho(t) \mid \rho(0) = \rho_k] = \bar{\rho}$, it consequently follows that $\sum_{k>k_0} \mu_k \mathrm{E}_0[\rho(t) \mid \rho(0) = \rho_k]$ increases with t.

[42] Any value $\underline{\theta} < \Theta$ would work.

Using first-order stochastic dominance once again shows that $\sum_{j,k>k_0} \hat{\mu}_{jk}\rho_j$ decreases in λ, and therefore that the price increases in λ.

Part (iii) follows immediately from direct computation. $\square$

A.1. Transaction costs

Suppose that agents may still trade exclusively at the Poisson times specified in Section 1, in addition to which they pay transaction costs proportional to the number of shares traded. It follows that, for any agent buying,

$$P + q = \mathrm{E}_0\left[\int_0^\infty e^{-rs}\kappa_1\bigl(\theta(s), \rho(s)\bigr)\,ds \,\Big|\, \theta(0) = \theta_k^b,\ \rho(0) = \rho_k\right], \tag{A.21}$$

while, for any seller,

$$P - q = \mathrm{E}_0\left[\int_0^\infty e^{-rs}\kappa_1\bigl(\theta(s), \rho(s)\bigr)\,ds \,\Big|\, \theta(0) = \theta_k^s,\ \rho(0) = \rho_k\right]. \tag{A.22}$$

Note that Eq. (A.21) holds for all agents of type (θ_j^b, ρ_j), while Eq. (A.22) holds for all agents of type (θ_j^s, ρ_j). Consequently, one of Eqs. (A.21)–(A.22) holds for any agent in the market at time t.

In order to write the demand schedules, let us denote the event of buying at t, or being indifferent to doing so, when of type k, by $(b,k)_t$. Define $(s,k)_t$ similarly. As in the case with immediate trade, a buyer currently considering acquiring a marginal unit either will save $P+q$ if he turns out to want to buy next time in the market, or will receive $P-q$ if he sells next time, so that

$$\begin{aligned} P + q = {} & \mathrm{E}_0\left[\int_0^\tau e^{-rs}\kappa_1\bigl(\theta_k^b, \rho(s)\bigr)\,ds \,\Big|\, (b,k)_0\right] \\ & + \mathrm{E}_0\bigl[e^{-r\tau}\bigl(P + q1_{(\text{buy at }\tau)} - q1_{(\text{sell at }\tau)}\bigr) \,\big|\, (b,k)_0\bigr], \end{aligned} \tag{A.23}$$

or, using the exponential arrival times and rewriting,

$$P = \frac{1}{r}\kappa_1\left(\theta_k^b, \sum_j \tilde{\mu}_{jk}\rho_j\right) - q - 2q\frac{r+\lambda}{r}\mathrm{E}_0\bigl[e^{-r\tau}1_{(\text{sell at }\tau)} \,\big|\, (b,k)_0\bigr]. \tag{A.24}$$

The analogous equation for a seller is

$$P = \frac{1}{r}\kappa_1\left(\theta_k^s, \sum_j \tilde{\mu}_{jk}\rho_j\right) + q + 2q\frac{r+\lambda}{r}\mathrm{E}_0\bigl[e^{-r\tau}1_{(\text{buy at }\tau)} \,\big|\, (s,k)_0\bigr] \tag{A.25}$$

and the resulting demand equations are

$$\begin{aligned} \theta_k^b = {} & \frac{1}{\gamma\sigma_D^2}\left(\frac{m_D}{r} - \left(P + q + 2q\frac{r+\lambda}{r}\mathrm{E}_0\bigl[e^{-r\tau}1_{(\text{sell at }\tau)} \,\big|\, (b,k)_0\bigr]\right)\right) \\ & - \frac{\sigma_\eta}{\sigma_D}\sum_j \tilde{\mu}_{jk}\rho_j, \end{aligned} \tag{A.26}$$

$$\theta_k^s = \frac{1}{\gamma\sigma_D^2}\left(\frac{m_D}{r} - \left(P - q - 2q\frac{r+\lambda}{r}\mathrm{E}_0\left[e^{-r\tau}1_{(\text{buy at }\tau)} \,\middle|\, (s,k)_0\right]\right)\right) - \frac{\sigma_\eta}{\sigma_D}\sum_j \tilde{\mu}_{jk}\rho_j. \tag{A.27}$$

Aggregating (40) and (41) yields the equilibrium price. To write it, let μ_{jk}^b denote the total mass of agents of type ρ_j who hold θ_k^b and define μ_{jk}^s analogously. Also, let $\mu^b = \sum_{j,k:\ j>k}\mu_{jk} + \sum_j \mu_{jj}^b$ and $\mu^s = 1 - \mu^b$. Thus, μ^b (μ^s) is the total mass of all agents who would be taking the position θ_j^b (θ_j^s) for some j if able to trade. The following holds.

Proposition 6. *There exists $\bar{q} > 0$ such that, for transaction costs $q \leqslant \bar{q}$, in steady state*

$$P = P^W - q\left(\mu^b - \mu^s\right) - 2q\frac{r+\lambda}{r}\mathrm{E}_0\left[e^{-r\tau}\left(1_{(\text{buy at 0, sell at }\tau)} - 1_{(\text{sell at 0, buy at }\tau)}\right)\right] \tag{A.28}$$

and θ_k^b and θ_k^s, given by (40) *and* (41)*, can be written as*

$$\theta_k^b = \theta_k^W + \frac{q}{\gamma\sigma_D\sigma_\eta}A^b, \tag{A.29}$$

$$\theta_k^s = \theta_k^W + \frac{q}{\gamma\sigma_D\sigma_\eta}A^s \tag{A.30}$$

with A^b and A^s independent of q and $A^b < A^s$.

Proof of Propositions 4 and 6. Note that Proposition 4 is a special case of Proposition 6, on which I can consequently concentrate.

First, I show that, for q low enough, all agents able to trade do so if their types do not correspond to their positions. With zero transaction costs, a strategy of never trading when of kind (ρ_j, θ_k) for some $j \neq k$ results in positive utility loss that is bounded below away from 0 regardless of λ. Since J is finite, by continuity, trading is optimal also for $q < \bar{q}$, if $\bar{q}$ is small enough.

It is immediate from (A.26)–(A.27) that $\theta_k^s > \theta_k^b$, or $A^s > A^b$. $\square$

A.2. Agent life cycle and exogenous transaction costs

In the context of Section 4.2.3, assume that agents have finite life spans. Specifically, suppose that every agent may have to leave the economy with intensity π. In this case, the agent has immediate access to the market, where he liquidates his position. The bequest function is defined as if the agent could only invest in the risk-free asset from then onwards:

$$\bar{V}(W) = -e^{-r\gamma W}. \tag{A.31}$$

Consequently, Eqs. (A.21)–(A.22) become

$$P + q = \mathrm{E}_0\left[\int_0^{\tau_\pi} e^{-rs}\kappa_1\big(\theta(s),\rho(s)\big)\,ds \,\middle|\, \rho(0),\ \text{buy at } 0\right] + \mathrm{E}_0\left[e^{-r\tau_\pi}\left(P - q1_{(\theta(\tau_\pi)>0)} + q1_{(\theta(\tau_\pi)<0)}\right) \,\middle|\, \rho(0),\ \text{buy at } 0\right], \tag{A.32}$$

$$P - q = \mathrm{E}_0\left[\int_0^{\tau_\pi} e^{-rs}\kappa_1\big(\theta(s), \rho(s)\big)\,ds \;\Big|\; \rho(0),\ \text{sell at } 0\right]$$
$$+ \mathrm{E}_0\big[e^{-r\tau_\pi}\big(P - q1_{(\theta(\tau_\pi)>0)} + q1_{(\theta(\tau_\pi)<0)}\big) \;\big|\; \rho(0), \text{sell at } 0\big], \tag{A.33}$$

where τ_π is the arrival time of exit.

To preserve stationarity, entry is assumed at the same rate as exit, types being drawn from the stationary distribution. When aggregating the pricing equations above, only agents choosing their position freely are considered: agents already owning positions and trading to different positions, and agents trading for the first time. Eqs. (A.32)–(A.33) do not hold for agents exiting the economy.

Let $\alpha^{k+} = \sum_{j<k}\alpha_{kj}$, $\alpha^{k-} = \sum_{j>k}\alpha_{kj}$, and $\alpha^k = \alpha^{k+} + \alpha^{k-}$. To simplify matters, consider the case when the exogenous transaction costs constitute the only trading friction, i.e., $\lambda \to \infty$. Let μ_k^b be the mass of agents of type k who bought the last time they traded, and similarly for μ_k^s. In addition, $\mu^i = \sum_k \mu_k^i$ for $i \in \{b, s\}$. Denoting by τ_α^k the first arrival time after time 0 of a type switch for an agent of type k, the pricing equations are

$$P + q = \mathrm{E}_0\left[\int_0^{\tau_\alpha^k\wedge\tau_\pi} e^{-rs}\kappa_1\big(\theta_k^b, \rho_k\big)\,ds\right]$$
$$+ \mathrm{E}_0\big[e^{-r\tau_\alpha^k}\big(P - q1_{(\rho(\tau_\pi)<\rho_k)} + q1_{(\rho(\tau_\pi)>\rho_k)}\big)1_{(\tau_\alpha^k<\tau_\pi)} \;\big|\; \rho(0) = \rho_k\big]$$
$$+ \mathrm{E}_0\big[e^{-r\tau_\pi}\big(P - q1_{(\theta(\tau_\pi)>0)} + q1_{(\theta(\tau_\pi)<0)}\big)1_{(\tau_\alpha^k>\tau_\pi)} \;\big|\; \rho(0) = \rho_k\big], \tag{A.34}$$

$$P - q = \mathrm{E}_0\left[\int_0^{\tau_\alpha^k\wedge\tau_\pi} e^{-rs}\kappa_1\big(\theta_k^s, \rho_k\big)\,ds\right]$$
$$+ \mathrm{E}_0\big[e^{-r\tau_\alpha^k}\big(P - q1_{(\rho(\tau_\pi)<\rho_k)} + q1_{(\rho(\tau_\pi)>\rho_k)}\big)1_{(\tau_\alpha^k<\tau_\pi)} \;\big|\; \rho(0) = \rho_k\big]$$
$$+ \mathrm{E}_0\big[e^{-r\tau_\pi}\big(P - q1_{(\theta(\tau_\pi)>0)} + q1_{(\theta(\tau_\pi)<0)}\big)1_{(\tau_\alpha^k>\tau_\pi)} \;\big|\; \rho(0) = \rho_k\big], \tag{A.35}$$

which yield

$$P + q = \frac{1}{r}\kappa_1\big(\theta_k^b, \rho_k\big) - 2q\frac{r + \alpha^k + \pi}{r}$$
$$\times \mathrm{E}_0\big[e^{-(r+\pi)\tau_\alpha^k}1_{(\rho(\tau_\pi)<\rho_k)} + e^{-(r+\alpha^k)\tau_\pi}1_{(\theta_k^b>0)} \;\big|\; \rho(0) = \rho_k\big]$$
$$= \frac{1}{r}\kappa_1\big(\theta_k^b, \rho_k\big) - \frac{2q}{r}\big(\alpha^{k+} + \pi 1_{(\theta_k^b>0)}\big), \tag{A.36}$$

$$P - q = \frac{1}{r}\kappa_1\big(\theta_k^s, \rho_k\big) + 2q\frac{r + \alpha^k + \pi}{r}$$
$$\times \mathrm{E}_0\big[e^{-(r+\pi)\tau_\alpha^k}1_{(\rho(\tau_\pi)>\rho_k)} + e^{-(r+\alpha^k)\tau_\pi}1_{(\theta_k^s<0)} \;\big|\; \rho(0) = \rho_k\big]$$
$$= \frac{1}{r}\kappa_1\big(\theta_k^s, \rho_k\big) + \frac{2q}{r}\big(\alpha^{k-} + \pi 1_{(\theta_k^s<0)}\big). \tag{A.37}$$

Aggregating Eqs. (A.36)–(A.37) over all agents in the economy results in

Proposition 7. *There exists $\bar{q} > 0$ such that, for transaction costs $q \leqslant \bar{q}$ and with entry and exit, the steady-state price equals*

$$P = P^W - q\left(\mu^b - \mu^s\right) - 2q\frac{\pi}{r}\left(1 - \sum_k\left(\mu_k^b 1_{(\theta_k^b<0)} + \mu_k^s 1_{(\theta_k^s<0)} + \mu_k 1_{(\theta_k^b<0)}\right)\right). \tag{A.38}$$

If there is no shorting in equilibrium, this equation simplifies to

$$P = P^W - q\left(\mu^b - \mu^s\right) - 2q\frac{\pi}{r}. \tag{A.39}$$

Proof. Eq. (A.38) owes to the zero-flow condition for the mass μ^b:

$$\sum_k \mu_k^b \alpha^{k+} + \pi\mu^b = \sum_k \mu_k^s \alpha^{k-} + \pi\sum_k \mu_k 1_{(\theta_k^b>0)}. \qquad \square$$

References

[1] Yakov Amihud, Haim Mendelson, Asset pricing and the bid-ask spread, J. Finan. Econ. 17 (1986) 223–249.

[2] Yakov Amihud, Haim Mendelson, Lasse Heje Pedersen, Liquidity and asset prices, Foundations and Trends in Finance, 2006.

[3] Adam Ashcraft, Darrell Duffie, Systemic illiquidity in the federal funds market, Amer. Econ. Rev. Pap. Proceed. 97 (2) (2007) 221–225.

[4] Jack Bao, Jun Pan, Excess volatility of corporated bonds, Working paper, MIT, 2008.

[5] John Y. Campbell, Albert S. Kyle, Smart money, noise trading and stock price behaviour, Rev. Econ. Stud. 60 (1993) 1–34.

[6] Susan Chaplinsky, David Haushalter, Financing under extreme uncertainty: Contract terms and returns to private investments in public equity, Working paper, University of Virginia, 2006.

[7] Raj Chetty, Adam Szeidl, Consumption commitments: Neoclassical foundations for habit formation, Working paper, University of California, Berkeley, 2004.

[8] George M. Constantinides, Capital market equilibrium with transaction costs, J. Polit. Economy 94 (1986) 842–862.

[9] Gene D'Avolio, The market for borrowing stock, J. Finan. Econ. 66 (2002) 271–306.

[10] Frank de Jong, Joost Driessen, Liquidity risk premia in corporate bond markets, Working paper, University of Amsterdam, 2006.

[11] Darrell Duffie, Yeneng Sun, Existence of independent random matching, Ann. Appl. Probab. 17 (2007) 386–419.

[12] Darrell Duffie, Nicolae Gârleanu, Lasse Heje Pedersen, Securties lending, shorting, and pricing, J. Finan. Econ. 66 (2002) 307–339.

[13] Darrell Duffie, Nicolae Gârleanu, Lasse Heje Pedersen, Over-the-counter markets, Econometrica 73 (2005) 1815–1847.

[14] Darrell Duffie, Nicolae Gârleanu, Lasse Heje Pedersen, Valuation in over-the-counter markets, Rev. Finan. Stud. 20 (6) (2007) 1865–1900.

[15] Xavier Gabaix, David Laibson, The 6d bias and the equity premium puzzle, in: Benjamin Bernanke, Kenneth Rogoff (Eds.), NBER Macroeconomics Annual, vol. 16, 2002, pp. 257–312.

[16] Nicolae Gârleanu, Lasse Heje Pedersen, Adverse selection and the required return, Rev. Finan. Stud. 17 (2004) 643–665.

[17] Nicolae Gârleanu, Lasse Heje Pedersen, Liquidity and risk management, Amer. Econ. Rev. Pap. Proceed. 97 (2007) 193–197.

[18] Christopher C. Geczy, David K. Musto, Adam V. Reed, Stocks are special too: An analysis of the equity lending market, J. Finan. Econ. 66 (2002) 241–269.

[19] Richard C. Green, Burton Hollifield, Norman Schürhoff, Financial intermediation and the costs of trading in an opaque market, Rev. Finan. Stud. 20 (2) (2007) 275–314.

[20] Sanford J. Grossman, Guy Laroque, Asset pricing and optimal portfolio choice in the presence of illiquid durable consumption goods, Econometrica 58 (1990) 25–51.

[21] J. Michael Harrison, David M. Kreps, Speculative investor behavior in a stock market with heterogeneous expectations, Quart. J. Econ. 92 (1978) 323–336.

[22] Ming Huang, Liquidity shocks and equilibrium liquidity premia, J. Econ. Theory 109 (2003) 104–129.
[23] Eleazer Klein, Fundamental pipes: Typical structures and transactions, in: Steven Dresner, Kurt Kim (Eds.), PIPEs: A Guide to Private Investments in Public Equity, Bloomberg Press, 2003, pp. 169–210.
[24] Ricardo Lagos, Guillaume Rocheteau, Search in asset markets, Working paper, NYU, 2006.
[25] Andrew W. Lo, Harry Mamaysky, Jiang Wang, Asset pries and trading volume under fixed transactions costs, J. Polit. Economy 112 (5) (2004) 1054–1090.
[26] Francis Longstaff, Financial claustrophobia: Asset pricing in illiquid markets, Amer. Econ. Rev., forthcoming.
[27] Mark Mitchell, Lasse Heje Pedersen, Todd Pulvino, Slow moving capital, Amer. Econ. Rev. Pap. Proceed. 97 (2) (2007) 215–220.
[28] Ricardo Reis, Inattentive consumers, J. Monet. Econ. 53 (8) (2006) 1761–1800.
[29] José Scheinkman, Wei Xiong, Overconfidence, short-sale constraints, and bubbles, J. Polit. Economy 111 (2003) 1183–1219.
[30] James R. Tannenbaum, Anna T. Pinedo, The law: Legal and regulatory framework, in: Steven Dresner, E. Kurt Kim (Eds.), PIPEs: A Guide to Private Investments in Public Equity, Bloomberg Press, 2003, pp. 109–168.
[31] Dimitri Vayanos, Transaction costs and asset prices: A dynamic equilibrium model, Rev. Finan. Stud. 11 (1998) 1–58.
[32] Dimitri Vayanos, Tan Wang, Search and endogenous concentration of liquidity in asset markets, J. Econ. Theory 136 (1) (2007) 66–104.
[33] Dimitri Vayanos, Pierre-Olivier Weill, A search-based theory of the on-the-run phenomenon, J. Finance 63 (2008) 1351–1389.
[34] Pierre-Olivier Weill, Liquidity premia in dynamic bargaining markets, J. Econ. Theory 140 (1) (2008) 66–96.
[35] Pierre-Olivier Weill, Leaning against the wind, Rev. Econ. Stud. 74 (2007) 1329–1354.