Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Economic Theory 109 (2003) 104–129
JOURNAL OF Economic Theory
http://www.elsevier.com/locate/jet

Liquidity shocks and equilibrium liquidity premia

Ming Huang*

Graduate School of Business, Stanford University, Stanford, CA 94305, USA

Received 3 December 1999; final version received 17 January 2002

Abstract

We study an equilibrium in which agents face surprise liquidity shocks and invest in liquid and illiquid riskless assets. The random holding horizon from liquidity shocks makes the return of the illiquid security risky. The equilibrium premium for such risk depends on the constraint that agents face when borrowing against future income; it is insignificant without borrowing constraint, but can be very high with borrowing constraint. Illiquidity, therefore, can have large effects on asset returns when agents face liquidity shocks and borrowing constraints. This result can help us understand why some securities have high liquidity premia, despite low turnover frequency.

JEL classification: G11; G12

Keywords: Liquidity shocks; Liquidity premium

1. Introduction

Trading financial securities requires transactions costs.[1] The impact of such costs on asset prices has been the subject of numerous empirical[2] and theoretical studies.[3]

*Fax: 650-725-6152.
E-mail address: mhuang@stanford.edu.

[1] A partial list of such costs includes bid-ask spread, market-impact costs, delay and search costs, and direct transactions fees (including brokerage commissions, exchange fees, and transactions taxes). See [6] for more details.

[2] See, for example, [4,5,8,9,33]. Other related empirical studies on effects of liquidity on asset returns [3,10,11,13,23,29].

[3] See, for example, [2,4,12,21,22,25,34–36]. Also see [1,24], for example, for effect of liquidity on asset prices in the context of asymmetric information, and see [18] for a model with endogenous liquidity correlated with productivity and investment.

0022-0531/03/$ - see front matter
doi:10.1016/S0022-0531(02)00039-X

Although we now have a good qualitative understanding of the effect of transactions costs on asset prices, existing theoretical studies have not enjoyed much success in explaining the empirically documented large impact of transactions costs on asset prices. Amihud and Mendelson [4] and Brennan and Subrahmanyam [9], for example, found that a significant part of cross-sectional variations in US stock returns can be attributed to difference in transactions costs. They show that the liquidity-premium coefficient (defined as the ratio between the differential of expected returns and the differential of (round-trip proportional) transactions costs of securities with similar cash flows but different transactions costs) ranges between 1.5 and 2.[4] In contrast, theoretical studies by Aiyagari and Gertler [2] (with investors trading riskless assets for consumption smoothening while facing income shocks), Constantinides [12] (with investors trading for portfolio rebalancing), Heaton and Lucas [22] (with investors trading riskless and risky assets to share income risk), Vayanos [35] and Vayanos and Vila [36] (with investors trading for life-cycle reason) all show that investors should drastically reduce their trading of the illiquid assets when facing transactions costs and demand only a very small liquidity premium, with the liquidity-premium coefficient typically less than 0.2 under reasonable parameter choices.

An important reason behind such discrepancy between theory and empirical findings is that existing theories on transactions costs—and in fact the asset pricing literature in general—have not yet given a quantitative explanation for the observed high market trading volume (see, for example, [17,25,26,28,30]). Since the liquidity premia of illiquid securities depend strongly on investors' holding horizon over which the transactions costs get amortized, theories that cannot account for the observed high trading volume (i.e., investors' high-frequency trading needs) inevitably face difficulty in explaining the observed market liquidity premia.

Moreover, even if one takes as given the observed trading frequency in the market, one may still find it difficult to reconcile between the market average holding horizon and the market liquidity premia. Consider, for example, the US stock market. With an average holding horizon of about two years, a simple method of amortizing the round-trip transactions costs over the holding horizon yields a liquidity-premium coefficient of 0.5, the inverse of the holding horizon. This number is still much smaller than those estimated by Amihud and Mendelson [4] and Brennan and Subrahmanyam [9].

Motivated by these observations, this paper develops a model of an economy in which investors face surprise liquidity shocks and invest in liquid and illiquid riskless assets. We find that the impact of transactions costs on asset returns in this economy is fairly large when investors are constrained from borrowing against future income.

Our approach departs from the existing theoretical literature on effects of transactions costs on asset prices by recognizing that some investors in the economy do not have control over the liquidation time of their security holdings, which could

[4] The fact that transactions costs have large impact on asset returns is also illustrated in a study by Silber [33] which shows that restricted ("letter") stocks that cannot be publicly traded for 2-years sell at an average discount of 35% below regular stocks.

come by surprise. A household may experience surprise liquidity shocks (such as a sudden drop in wealth or a surprise consumption need) and, facing a borrowing constraint, must liquidate its illiquid assets. Firms have surprise investment opportunities, but face an imperfect and costly external capital market, and therefore hold reserves to be liquidated when such opportunities arrive. Professional traders hold reserves so as to take advantage of certain surprise news event by creating new portfolio positions. Fund managers worry about a sudden increase in withdrawal, resulting in a need to liquidate some illiquid assets. In all these situations, it may be better to model the motivation for trade by a random arrival time of liquidation. In contrast, previous theoretical studies have assumed that the holding period for securities is either known or controlled by the investor and predictable.

Our analysis shows that, for investors who face liquidity shocks, the required liquidity premium depends not only on the expected holding horizon, which determines investors' average frequency of trading needs, but also on the surprise (and random) nature of such holding horizon. Although the randomness of the holding horizon only has a small effect on the equilibrium liquidity premium in an economy without any borrowing constraint, it has a large effect on liquidity premium under the more realistic assumption that investors are constrained from borrowing against future income.

The intuition behind our results can be illustrated by comparing the illiquid asset with the (otherwise identical) liquid asset. In equilibrium, the illiquid asset should generate higher expected (pre-transactions-costs) return to compensate its holders for illiquidity. In a non-stochastic environment, one can use the present-discounted-value (PDV) approach, in which the price of the illiquid asset would simply be the price of the liquid asset adjusted for the present discounted value of transactions costs (with the riskless rate as the discount rate), such that the *net-of-transactions-costs* holding-horizon return for the illiquid asset is the riskless rate. In the economy studied here, however, this approach no longer works because of liquidity shocks.[5] Specifically, the (net-of-transactions-costs) return of the illiquid asset is now risky—it is low (and even possibly negative) for agents who experience liquidity shocks soon after buying it, but high for agents who experience late liquidity shocks.

The question is whether such risk of the illiquid-asset return is priced in equilibrium. If agents can borrow against future income at any time, then their consumption levels at liquidity shocks of different times are smooth and they do not require significant risk premium for holding the illiquid asset. In this case, agents demand roughly the same liquidity premium as those with a fixed holding horizon that is equal to the expected arrival time of the liquidity shock. On the other hand, under the more reasonable assumption that agents accumulate income over time and cannot borrow against future income stream, the unlucky agents who suffer an early liquidity shock also have had less time to accumulate income in preparation for the shock. They therefore require a risk premium for holding the illiquid asset. This

[5] We assume that there exists no mechanism that allows for insurance against the liquidity shock risk because of moral hazard or adverse selection problems.

leads to a higher equilibrium liquidity premium which increases with agents' level of risk aversion.

Our results of potentially large liquidity premia can be helpful in attempts to bridge the quantitative gap of liquidity premia between theoretical calculations and empirical estimates. In particular, they can be helpful in understanding why some securities have high liquidity premia, despite their relatively low turnover frequency.

Our study also contributes to the vast theoretical literature on the impact of transactions costs on investors' portfolio policy[6] by analyzing such impact for investors who face surprise liquidity shocks.

The model is cast in a continuous-time overlapping-generations economy with two consol bonds, one liquid and one illiquid. In order to focus on the impact of transactions costs and liquidity shocks, we take both consols to be riskless. Each pays dividends at a constant rate. In order to model the surprise nature of investors' need to liquidate, the paper uses a Poisson arrival time to (exogenously) represent the arrival of a liquidity shock. Each investor in the economy derives his utility from "consumption" at the time of liquidity shock.

In Section 2, we specify the model and the typical investor's control problem. In Section 3, we consider investors who are endowed up front with a lump-sum of wealth upon their "entry" into the economy, and interpret this case to be inclusive of cases with intertemporal income but no borrowing constraints (because investors can monetarize their future income anytime). Section 4 treats investors who face borrowing constraint and have to rely on past income and accumulated wealth to deal with liquidity shock. Section 5 concludes. All proofs are in the appendix.

2. The model

We consider a continuous-time overlapping generations economy with a continuum of agents having a total mass of 1.[7] At any time, each agent faces a constant probability per unit time, λ, of experiencing a liquidity shock which, upon arrival, will force him to liquidate his securities and exit the economy.[8] The arrivals of the liquidity shocks for all existing agents are independent Poisson times. New agents enter the economy at a constant rate of λ per unit time. Because of the law of large numbers, agents exit at a rate of λ per unit of time, leaving the economy with a constant mass of agents.

[6] See, for example, [12,14–16,19,27,31,32].

[7] Without loss of generality, our set of agents is the unit interval. Independence across the set of agents, and the effect of law of large numbers, are achieved by the usual devices, as in [7] or [20].

[8] The assumption that agents exit the economy upon a liquidity shock is made for technical simplicity. The qualitative results in this paper do not rely on the assumption.

2.1. Financial structure

There is a single consumption good which also serves as the numeraire. There are two financial securities. Both are riskless consol bonds, that is, perpetuities that pay a constant flow of consumption dividends at a rate of d per unit of time. One consol is liquid, with a price of p per share, and no transactions costs are required for trading it. The other is illiquid, with a price of P per share, and agents who buy or sell x shares of it pay a proportional transactions cost totaling $\varepsilon x P$.[9] No short sales are allowed for either consol.[10] The total supply of consols, both liquid and illiquid, is normalized to 1, so that the aggregate dividend rate (from both) is d per unit of time. The fractional supply of the liquid consol is $k \in (0, 1)$. (That is, the total dividend rate from the liquid consol is kd per unit of time.)

We will construct equilibria in which the prices of the liquid and the illiquid consols are, respectively, constant at p and P. The dividend flow rate (per unit of numeraire) for the liquid and illiquid consols are, respectively, $r = d/p$ and $R = d/P$. We assume that $R \geqslant (1 + \varepsilon)r$, for otherwise the liquid consol dominates the illiquid consol. Our goal is to find and characterize the optimal investment policy of investors and the equilibrium liquidity premium, $R - r$.

2.2. Preferences

Each agent has utility $\mathbf{E}[U_T(C_T)]$ for terminal consumption, C_T, at the random arrival time T of that agent's liquidity shock. Each agent has constant time-preference rate $\beta \geqslant 0$ and CRRA utility with a constant relative risk-aversion γ. That is,

$$U_T(C_T) = \begin{cases} e^{-\beta T} \frac{C_T^{1-\gamma}}{1-\gamma} & \text{if } \gamma \neq 1, \\ e^{-\beta T} \log(C_T) & \text{if } \gamma = 1. \end{cases} \tag{1}$$

2.3. An agent's control problem

We consider a particular agent who enters into the economy at time 0 with an endowment W_0 of initial wealth, and (exogenously specified) deterministic income at the rate $y_t \geqslant 0$ at time t, for some y such that $\int_0^t y_s \, ds < \infty$ for all t. The agent invests current wealth, endowed income, and dividends in the liquid and illiquid consols. Let the processes a and A denote the agent's holdings of the liquid and illiquid consols, respectively. The short-sale constraint implies that $a \geqslant 0$ and $A \geqslant 0$. Because, in the

[9] The analytic tractability, as well as the economic conclusions in the paper, does not rely on the assumption that buyers and sellers equally share the total transactions costs of $2\varepsilon x P$.

[10] By imposing the prohibition against short sales, we are effectively assuming, reasonably, that short sales are so costly that the price differential between the two securities is not at the upper bound implied by the no-arbitrage condition, and is instead determined by investors' equilibrium demand for the securities. We are interested here in the determinants of the securities' equilibrium return differential within the no-arbitrage bounds.

equilibrium that we examine, there is no payoff-relevant information before the agent's liquidity shock at time T, we may without loss of generality take both a and A to be deterministic. Moreover, because of the existence of non-zero transactions costs for the illiquid consol, the agent's cumulative volume of trade in both the liquid and illiquid consols is (without loss of generality) finite up to any time t, for otherwise an infinite amount of wealth would be lost in transactions costs. We therefore can take both a and A to be of finite variation.

Taking the expectation of $U_T(C_T)$ over the Poisson arrival time T, the agent thus faces the problem

$$\max_{(a,A)} \mathbf{E}[U_T(C_T)], \tag{2}$$

where T has the exponential distribution with parameter λ, subject to the short-sale constraint and portfolio dynamics:[11]

$$a_t \geqslant 0; \quad A_t \geqslant 0, \tag{3}$$

$$y_t\,dt = da_t + dA_t + \varepsilon dL_t^A + \varepsilon dD_t^A - ra_t\,dt - RA_t\,dt, \tag{4}$$

$$W_0 = a_0 + (1+\varepsilon)A_0, \tag{5}$$

$$C_t = a_t + (1-\varepsilon)A_t, \tag{6}$$

where L^A and D^A denote the cumulative value of the illiquid consol purchased and sold, respectively, and therefore $A = L^A - D^A$. Because of (6), C is deterministic, and we may therefore re-write our problem as

$$\max_{(a,A)} \int_0^\infty U_t(C_t)e^{-\lambda t}\lambda\,dt,$$

subject to (3)–(6). To ensure finiteness of the investor's utility, we assume that $\beta + \lambda - \frac{R}{1+\varepsilon} > 0$.

3. Optimal policy and equilibrium: initial wealth only

In this section, we consider the extreme case in which each agent's endowment is a lump-sum of wealth at time zero. That is, $W_0 > 0$ and $y_t = 0$ for all t. Agents are thus well prepared for liquidity shocks upon their "entries" into the economy. We can also think of this setup as a case in which agents do have income over time but can monetarize their future income stream by borrowing against future income.

We first study the optimal investment policy of a single agent entering at time zero. We later solve for the equilibrium and calculate the equilibrium liquidity premium.

[11] The symbol d in (4) denotes differential, not the dividend flow rate.

3.1. An agent's optimal policy

The problem here is to find the policy (a, A) that solves the maximization problem (2)–(6), given a specification of preferences (γ, β), the arrival intensity (or "hazard rate") λ of liquidity shocks, and the financial market parameters (ε, r, and R). With $y = 0$, this control problem is simplified by its stationarity and admits an exact solution.

We first derive heuristic conditions for the optimal policy. These conditions will help us identify a candidate optimal policy. We will, of course, prove later that the identified candidate policy is indeed optimal.

Depending on the liquidity premium, an agent may choose to hold the liquid consol only, the illiquid consol only, or a mixed portfolio of the liquid and illiquid consols.

Consider a candidate optimal policy of always holding, and reinvesting in, the illiquid consol only. Consider an alternate policy that deviates from the candidate optimal policy from time t to time $t + \Delta t$, by investing one unit (to be thought of as a "small" amount) in the liquid consol at time t and then converting the resulting liquid consol holding back to the illiquid consol at time $t + \Delta t$. By adopting such a deviation, the agent's consumption increases[12] by $1 - \frac{1-\varepsilon}{1+\varepsilon} + O(\Delta t)$, from the saving of the round-trip transactions costs, if the liquidity shock arrives within $(t, t + \Delta t)$.[13] If the liquidity shock arrives at $s > t + \Delta t$, then the agent's consumption decreases by

$$\frac{1-\varepsilon}{1+\varepsilon}\left(\frac{R}{1+\varepsilon} - r\right)\exp\left[\frac{R}{1+\varepsilon}(s-t)\right]\Delta t + o(\Delta t), \tag{7}$$

due to the lower dividends from the liquid consol holding within $(t, t + \Delta t)$.[14] For the candidate policy to be optimal, the agent's conditional expected utility must decrease, to the first order of Δt, when he adopts the alternate policy. That is, for any $t \in [0, \infty)$,

$$\mathscr{G}_t(\varepsilon, r, R) \equiv 2\varepsilon U_t'(C_t) - \frac{1-\varepsilon}{1+\varepsilon}[R - (1+\varepsilon)r] \int_t^\infty U_s'(C_s) e^{(\frac{R}{1+\varepsilon} - \lambda)(s-t)} \leqslant 0. \tag{8}$$

We can think of (8) as a first-order condition for optimality of (a, A).

---

[12] As usual, one says that "$f(x) = O(x)$" if $\lim_{x\to 0}|f(x)/x| \leqslant C$ for some $C > 0$, and that "$f(x) = o(x)$" if $\lim_{x\to 0}|f(x)/x| = 0$.

[13] Suppose the liquidity shock arrives at $s \in (t, t + \Delta t)$. One unit invested at t according to the candidate optimal policy becomes $\frac{1}{1+\varepsilon}$ shares of the illiquid consol at t (after transactions costs). This will grow to $\frac{1}{1+\varepsilon}\exp[\frac{R}{1+\varepsilon}(s-t)]$ shares at s, which generates $\frac{1-\varepsilon}{1+\varepsilon}\exp[\frac{R}{1+\varepsilon}(s-t)] = \frac{1-\varepsilon}{1+\varepsilon} + O(\Delta t)$ for $s \in (t, t+\Delta t)$ in revenue after transactions costs. Under the alternate policy, one unit invested at t grows to $e^{r(s-t)} = 1 + O(\Delta t)$ at s. The difference is $1 - \frac{1-\varepsilon}{1+\varepsilon} + O(\Delta t)$.

[14] In this case, one unit invested at t according to the candidate optimal policy pays off $\frac{1-\varepsilon}{1+\varepsilon}\exp[\frac{R}{1+\varepsilon}(s-t)]$ at time s (see footnote 13). Under the alternate policy, one unit invested at t becomes $\frac{1}{1+\varepsilon}\exp[r\Delta t]$ shares of illiquid consols after conversion at $t + \Delta t$, which becomes $\frac{1}{1+\varepsilon}\exp[r\Delta t]\exp[\frac{R}{1+\varepsilon}(s - t - \Delta t)]$ shares of illiquid consols at time s, with a liquidation revenue of $\frac{1-\varepsilon}{1+\varepsilon}\exp[r\Delta t]\exp[\frac{R}{1+\varepsilon}(s - t - \Delta t)]$ at time s. The difference gives (7).

Now consider a candidate optimal policy of always holding, and reinvesting in, the liquid consol only. Consider an alternate policy that deviates from the candidate policy by investing one unit in the illiquid consol at time t and reinvesting all of its dividends in the liquid consol.[15] By adopting such a deviation, in the event of a liquidity shock at some time $s>t$, the agent's consumption changes by[16]

$$\frac{1-\varepsilon}{1+\varepsilon}+\frac{1}{1+\varepsilon}\int_t^s Re^{r(s-u)}\,du-e^{r(s-t)}=\frac{1}{1+\varepsilon}\left[\frac{R-(1+\varepsilon)r}{r}[e^{r(s-t)}-1]-2\varepsilon\right]. \tag{9}$$

At the optimum, such a deviation decreases the agent's expected utility, which implies, for any $t\in[0,\infty)$, the associated first-order condition of optimality for the candidate policy is

$$\mathscr{H}_t\equiv\int_t^\infty U_s'(C_s)\left[\frac{R-(1+\varepsilon)r}{r}[e^{r(s-t)}-1]-2\varepsilon\right]e^{-\lambda(s-t)}\lambda\,ds\leqslant 0. \tag{10}$$

Finally, if the candidate optimal policy is to invest in both the liquid and the illiquid consols at all times, the associated first-order condition is

$$\mathscr{G}_t(\varepsilon,r,R)=\mathscr{H}_t=0,\quad t\in[0,\infty). \tag{11}$$

It is worth noting that the above derivation highlights the fact that the return of the illiquid consol is risky—it is low in the case of an early liquidity shock and high for a late liquidity shock (see, for example, (9)), and that in equilibrium, agents' degree of aversion toward such risk is determined by their levels of consumption at liquidity shocks at different times (see, for example, (11)).

One can construct an agent's optimal policy (a,A) from (8) to (11). According to this optimal policy, an agent invests all of his wealth and dividends in the liquid consol if the liquidity premium is too low; invests all of his wealth and dividends in the illiquid consol if the liquidity premium is sufficiently high; and maintains a constant proportion of the liquid and illiquid consols provided the liquidity premium is within an appropriate range. This result is given in the following proposition.

Proposition 1. *Suppose $W_0>0$ and $y=0$. The optimal policy (a,A) is given by*

$$\frac{a_t}{a_t+A_t}=\begin{cases}1 & \text{if } R\leqslant R^*,\\ \theta(R) & \text{if } R\in(R^*,R^{**}),\\ 0 & \text{if } R\geqslant R^{**},\end{cases} \tag{12}$$

[15] We do not consider a local deviation of policy here because it is never optimal to sell the illiquid consol before a liquidity shock.

[16] One unit invested at t according to the candidate policy grows to $e^{r(s-t)}$ at s. One unit invested at t according to the alternate policy generates $\frac{1}{1+\varepsilon}$ shares of the illiquid consol (which pays off $\frac{1-\varepsilon}{1+\varepsilon}$ at s), plus the dividends rolled over into the liquid consol between time t and s, $\frac{1}{1+\varepsilon}\int_t^s Re^{r(s-u)}\,du$. The difference between the two approaches gives (9).

where

$$R^* = r + 2\varepsilon\Big(\beta + \lambda + \gamma r - \frac{r}{2}\Big), \tag{13}$$

$$R^{**} = r + \frac{2\varepsilon(\beta + \lambda + \frac{\gamma r}{1+\varepsilon} - \frac{r}{2})}{1 - \frac{2\varepsilon}{1+\varepsilon}\gamma}, \tag{14}$$

and, for $R^* < R < R^{**}$,

$$\theta(R) = \frac{1 - \varphi(R)}{1 - \frac{\varepsilon}{1+\varepsilon}\varphi(R)} \tag{15}$$

with

$$\varphi(R) \equiv \frac{R - R^*}{R - (1+\varepsilon)r}\left[\frac{R^{**} - R^*}{R^{**} - (1+\varepsilon)r}\right]^{-1}. \tag{16}$$

The optimal wealth at time t, *assuming no liquidity shock by that time, is*

$$a_t + A_t = \begin{cases} W_0 \exp(rt) & \text{if } R \leqslant R^*, \\ \frac{W_0}{\theta(R)+(1+\varepsilon)(1-\theta(R))} \exp\left[\frac{\theta(R)r+(1-\theta(R))R}{\theta(R)+(1+\varepsilon)(1-\theta(R))}\, t\right] & \text{if } R^* < R < R^{**}, \\ \frac{W_0}{1+\varepsilon} \exp\left(\frac{R}{1+\varepsilon}\, t\right) & \text{if } R \geqslant R^{**}. \end{cases} \tag{17}$$

The agent holds the liquid consol exclusively if the liquidity premium is lower than $R^* - r$, and holds the illiquid consol exclusively if the liquidity premium is higher than $R^{**} - r$. If the liquidity premium falls between $(R^* - r, R^{**} - r)$, (15) and (16) show that the agent holds constant non-zero proportions of each consols. The illiquid-consol proportion, $1 - \theta(R)$, increases with the return R of the illiquid consol.

Although the agent has, at any time, a constant expected horizon $1/\lambda$ for the liquidity shock, his optimal portfolio policy differs from that of an agent having a fixed investment horizon. The latter is willing to mix the liquid and illiquid consols in his holdings only if the liquidity premium is such that the rates of return for the two consols, net of transactions costs, are exactly the same. Facing liquidity shocks at an uncertain horizon, an agent is willing to hold a mixed portfolio of the liquid and illiquid consols for any liquidity premium within the range $(R^* - r, R^{**} - r)$. We call this range the *mixed-portfolio range*, the size of which is

$$R^{**} - R^* = \frac{4\varepsilon^2\gamma}{1 + \varepsilon - 2\varepsilon\gamma}\,(\beta + \lambda - r + \gamma r). \tag{18}$$

This difference between a fixed and a random investment horizon can be understood as follows. If an agent is risk-averse and faces an uncertain investment horizon, the *net* return generated by the illiquid consol is risky as it depends on the random holding horizon. It is lower than that of the liquid consol in the case of an early shock, and is higher in the case of a late shock. A risk-averse agent is therefore willing to hold a mixed portfolio of the liquid and illiquid consols, given a liquidity premium within an appropriate range, in order to balance the benefit of liquidity (from the liquid consol), in case of an early shock, and the advantage of higher return (from the illiquid consol) in case of a late shock. As expected, this range shrinks to a single liquidity premium if the agent is risk-neutral. That is, if $\gamma = 0$, then, from (18), $R^* = R^{**}$.

If transactions costs are small, the size of the mixed-portfolio range is also small because, as can be seen from (17), $R^{**} - R^*$ is of second order in ε. This result is intuitive since the liquid and illiquid consols are closer substitutes as ε gets smaller.

3.2. Equilibrium for $W_0 > 0$ and $y = 0$

We now consider market equilibrium for the continuum, $[0, 1]$, of agents. The parameters $(W_0, \lambda, \beta, \varepsilon, k)$ define an economy. Assuming constant r and R, which we show below in equilibrium, all agents have the same problem (2)–(6), except for the starting time. An equilibrium for such an economy is a pair (r, R) of rates of return such that the aggregate demand for each consol at any point of time is equal to the aggregate supply of that consol. That is, let $a_{i,t}$ and $A_{i,t}$ denote the demand, in absolute value, by current agent $i \in [0, 1]$ for the liquid and illiquid consols at time t, respectively. Then market clearing means that[17]

$$\int_{[0,1]} a_{i,t}\, di = kp \quad \text{and} \quad \int_{[0,1]} A_{i,t}\, di = (1-k)P. \tag{19}$$

Under our conjecture that equilibrium prices are constant, as shown in Proposition 1, the proportions of the liquid and illiquid consols in each agent's optimal portfolio do not depend on time, and are thus the same for all agents in the economy. Given the finite supply of both consols, every agent must be an marginal investor of the illiquid consol in equilibrium. The equilibrium rate of return for the illiquid consol, R, must therefore be in (R^*, R^{**}).

The equilibrium rates of return, r and R, are determined by the market clearing conditions of (19). We can re-express these two market-clearing conditions as follows. The optimal proportions of the liquid and illiquid consols for each agent are the same as the proportion of these consols in the aggregate supply. That is,

$$\theta(R) = \frac{kp}{kp + (1-k)P} = \frac{kR}{kR + (1-k)r}. \tag{20}$$

The second condition is that the total market wealth should be the same as the total supply of consols. Assuming that the age distribution of the population is at its steady state at time 0, the law of large numbers implies that this distribution always has the exponential distribution with parameter λ (almost surely), or, in other words, that the density of ages is $e^{-\lambda t}$ at age t. We thus have

$$\begin{aligned} k\frac{d}{r} + (1-k)\frac{d}{R} &= \int_0^\infty (a_t + A_t) e^{-\lambda t} \lambda\, dt \\ &= \frac{\lambda W_0}{(\lambda - r)\theta(R) + [(1+\varepsilon)\lambda - R][1 - \theta(R)]}. \end{aligned} \tag{21}$$

[17] We mean (19) to hold almost surely, in the sense of the law of large numbers, as discussed in footnote 7.

Eqs. (20) and (21), combined with (15) and (16), determine the equilibrium rates of return. It can be shown that there exists some (r, R) that satisfies (15), (16), (20), and (21). The conjectured equilibrium is therefore valid.

The absolute magnitude of the rates of return r and R are relatively uninteresting because they are determined to a large extent by the exogenously specified aggregate dividend flow rate, d, and by the investors' aggregate endowment, W_0. The liquidity premium $R - r$ does not, however, depend on d and W_0 given r; it depends only on the *relative* supply of the two consols, investors' preferences, and the frequency of liquidity shocks. We thus take r as given and characterize the liquidity premium.

Given r, the solution for R is given by (16) and (20). Without solving for R explicitly, we note that, since R must be bounded by R^* and R^{**},

$$2\left(\beta + \lambda + \gamma r - \frac{r}{2}\right) < \frac{R - r}{\varepsilon} < \frac{2}{1 - \frac{2\varepsilon}{1+\varepsilon}\gamma}\left(\beta + \lambda + \frac{\gamma r}{1+\varepsilon} - \frac{r}{2}\right).$$

If all investors are risk-neutral, then the equilibrium liquidity-premium coefficient is

$$R - r = 2\varepsilon(\beta + \lambda - r/2).$$

The liquidity premium is higher if investors are risk-averse. The difference, to the first order of ε, is $2\varepsilon\gamma r$, which is small for typical parameters. For example, if $\varepsilon = 1\%$ and $r = 4\%$, then an increase in γ from 0 to 1 increases the equilibrium liquidity premium $R - r$ by about $2\varepsilon\gamma r = 8$ basis points.

The small impact of risk-aversion on the liquidity premium can be understood as follows. If an agent has an up-front endowment at time 0 and no income over time (or if he has intertemporal income but faces no borrowing constraints), then his consumption at an early liquidity shock differs from that at a late liquidity shock only by the amount of interest income, which is small. Such agents only demand a very small risk premium for holding the illiquid asset. In fact, for reasonable parameters (with β and r much smaller than λ), the liquidity premium is close to $2\varepsilon\lambda$, which is roughly the liquidity premium that one would obtain if one uses the PDV approach.

The impact on the liquidity premium of the relative supplies of the liquid and illiquid consols, k and $1 - k$, is only of second order. This is because, with only a lump-sum initial wealth endowment, all investors are marginal investors in the illiquid consol. A change in the relative supplies of the liquid and illiquid consols does not change the marginal investor.

More realistically, however, we expect that investors in the economy have different degrees of readiness for liquidity shocks and have an incentive to accumulate over time a wealth reserve in preparation for the shock. Moreover, agents are constrained from borrowing against future income. In the next section, we show that the impact of liquidity shocks on the liquidity premium can be quite large after we take these factors into account.

4. Optimal policy and equilibrium: intertemporal income flow

In this section, we assume that agents have steady income over time but cannot borrow against their future income. We introduce heterogeneity of agents' readiness for liquidity shocks by assuming that agents enter the economy with no initial endowment, but with an (exogenously specified) deterministic income flow. That is, we now take the case of $W_0 = 0$ and $y > 0$. We will show that, in equilibrium, agents demand risk premium for holding the illiquid asset, and the liquidity premium can be quite large.

As before, we first study a given agent's problem, then characterize market equilibrium.

4.1. Necessary and sufficient conditions for the optimal policy

We first derive the heuristic conditions for the optimal policy based on intuition and simple perturbation arguments, and then prove that these conditions are both necessary and sufficient for optimality.

When agents face a random holding horizon due to liquidity shocks, the net (annualized) return on the illiquid asset is risky—it is low in the case of an early liquidity shock, but high for a late liquidity shock. For the setup considered in this section, an agent builds up wealth reserve over time and is constrained from borrowing against his future income. He, therefore, has less to consume in the case of an early liquidity shock than in a late liquidity shock. Because of this undesirable positive correlation between his consumption level and the return on the illiquid asset, the agent is reluctant to hold the illiquid asset without a sufficient level of liquidity premium.

Moreover, for a given level of liquidity premium, a "younger" agent will find the illiquid asset more undesirable than an "older" agent. This is because the younger agent has accumulated less wealth reserve, and the asymmetry between his consumption level at an early liquidity shock and that in a late liquidity shock is more severe than that of an older agent. Therefore, for a given liquidity premium, we expect that an agent is more likely to choose the liquid consol early in his "life." Depending on the liquidity premium, the agent may switch to the illiquid consol after having built up "sufficient" reserves. So we search for the agent's optimal policy from within the class of policies with the following property: There exist t^* and t^{**}, with $0 \leqslant t^* < t^{**} \leqslant \infty$, such that the agent holds only the liquid consol for $t \leqslant t^*$, holds the liquid consol and accumulates the illiquid consol for $t \in (t^*, t^{**})$, and holds only the illiquid consol for $t \geqslant t^{**}$. That is,

$$\begin{cases} A_t = 0, & t \leqslant t^*, \\ a_t > 0, A_t > 0, & t \in (t^*, t^{**}), \\ A_s - A_t > 0, & t^* \leqslant t < s \leqslant t^{**}, \\ a_t = 0, & t \geqslant t^{**}. \end{cases} \tag{22}$$

Note that, between t^* and t^{**}, the agent holds a mixed portfolio in order to achieve an optimal balance between their need for short-term liquidity and long-term yield, which is the same reason behind the mixed-portfolio policy discussed in Section 3.1.

We now derive heuristic conditions of optimality for such a policy. Consider an agent with an optimal policy (a, A) satisfying (22). At any $t \geqslant t^{**}$, the agent derives less expected utility if he deviates from the above policy by investing one unit in the liquid consol and then converting the resulting liquid consol investment back to the illiquid consol at time $t + \Delta t$. This implies, from the same argument in Section 3.1 leading to (8), the first-order condition

$$\mathscr{G}_t(\varepsilon, r, R) \leqslant 0, \qquad t \geqslant t^{**}. \tag{23}$$

At any $t \in (t^*, t^{**})$, the agent is marginally indifferent between investing one unit in the illiquid consol, and investing one unit in the liquid consol and then converting it back to the illiquid consol at time $t + \Delta t \in (t^*, t^{**})$. This implies that

$$\mathscr{G}_t(\varepsilon, r, R) = 0, \quad t \in (t^*, t^{**}). \tag{24}$$

Now consider any time $t < t^*$. The agent derives less expected utility if he deviates from the above policy by investing one unit in the illiquid consol at t, rolling over its dividends into the liquid consol until t^*, and then converting all liquid holdings at t^* and all dividend payoffs after t^* into the illiquid consol. With this deviation, the agent's consumption upon liquidity shock at time s changes by[18]

$$\Delta C_s = \begin{cases} \dfrac{1}{1+\varepsilon}\left[\dfrac{R-(1+\varepsilon)r}{r}\left[e^{r(s-t)}-1\right]-2\varepsilon\right] & \text{if } s \leqslant t^*, \\ \dfrac{1-\varepsilon}{(1+\varepsilon)^2}\,\dfrac{R-(1+\varepsilon)r}{r}[e^{r(t^*-t)}-1]e^{\frac{R}{1+\varepsilon}(s-t)} & \text{if } s > t^*. \end{cases} \tag{25}$$

Optimality implies, for any $t < t^*$, the first-order condition

$$\begin{aligned} \mathscr{K}_t \equiv & \int_t^{t^*} U_s'(C_s)\left[\frac{R-(1+\varepsilon)r}{r}[e^{r(s-t)}-1]-2\varepsilon\right]e^{-\lambda(s-t)}\lambda\, ds \\ & + \frac{1-\varepsilon}{1+\varepsilon}\frac{R-(1+\varepsilon)r}{r}[e^{r(t^*-t)}-1]\int_{t^*}^{\infty} U_s'(C_s)e^{\frac{R}{1+\varepsilon}(s-t^*)}e^{-\lambda(s-t)}\lambda\, ds \leqslant 0. \end{aligned} \tag{26}$$

The heuristic first-order conditions (23)–(26) are necessary for optimality of a policy (a, A) satisfying (22). It turns out that conditions (22)–(26) are also sufficient conditions for optimality. The result is stated below and proved in the appendix.

[18] If the proposed candidate policy is optimal, then it is marginally optimal to invest one unit in the liquid consol at t, roll over dividends into the liquid consol before t^*, and invest only in the illiquid consol after t^*. Under such a proposed policy, one unit invested at t grows to $e^{r(s-t)}$ at any time $s \leqslant t^*$, becomes $\frac{1}{1+\varepsilon}e^{r(t^*-t)}$ shares of the illiquid consols at t^*, and pays off $\frac{1-\varepsilon}{1+\varepsilon}e^{r(t^*-t)}\exp[\frac{R}{1+\varepsilon}(s-t^*)]$ at any time $s > t^*$. Under the alternate policy, however, one unit invested at t becomes, at any time $s \leqslant t^*$, $\frac{1}{1+\varepsilon}$ shares of the illiquid consol (which pays off $\frac{1-\varepsilon}{1+\varepsilon}$), plus, from the dividends rolled over into the liquid consol between time t and s, $\frac{1}{1+\varepsilon}\int_t^s Re^{r(s-u)}\,du$. At $s > t^*$, the total illiquid consol holding is $\frac{1}{1+\varepsilon}\exp[\frac{R}{1+\varepsilon}(s-t^*)] + \frac{R}{(1+\varepsilon)^2 r}[e^{r(t^*-t)} - 1]\exp[\frac{R}{1+\varepsilon}(s-t^*)]$, which, when multiplied by $1-\varepsilon$, gives the liquidation revenue at time $s > t^*$. The difference between the two policies gives Eq. (25).

Lemma 1. *A policy* (a, A) *of the form of* (22) *is optimal if and only if it satisfies* (23)–(26).

For the general case with $y > 0$, the necessary and sufficient conditions for the optimal policy, (23)–(26), are difficult to solve analytically since they are a continuum collection of integral equations.[19] Fortunately, for special cases with "small" transactions costs (in a sense to be made precise), we can solve (23)–(26) to obtain the optimal policy. This solution, presented in Section 4.2, allows us to solve for, in Section 4.3, the equilibrium for cases with small transactions costs. The equilibrium liquidity-premium coefficient, which represents the impact of transactions costs on asset prices after all, can still be analyzed and calculated, which we do in Section 4.4. Luckily, the limited analytical tractability still allows us to obtain most of the economics in our model.

4.2. The optimal policy for small transactions costs

For analytical tractability, we consider the limit of an agent's optimal policy as $\varepsilon \to 0$. In this limit, an agent's control problem is interesting only if $R - r \to 0$ as well. Therefore we have to consider a sequence of economies defined by a sequence $\{(\varepsilon_n, r_n, R_n)\}$. We first solve for an agent's limiting optimal control solution, as $\varepsilon_n \to 0$, $r_n \to r > 0$, and $\frac{R_n - r}{2\varepsilon_n} \to K > 0$. This solution depends on the liquidity-premium coefficient K.

As the agent builds up a wealth reserve, he becomes less concerned about the riskiness of the returns of the illiquid asset. Intuitively, we expect that his optimal policy takes the form of (22). For the nth economy, defined by $(\varepsilon_n, r_n, R_n)$, let $C_t^n = a_t^n + (1 - \varepsilon)A_t^n$ denote the agent's consumption, according to his optimal policy, at time t. As $n \to \infty$, for any $t \geqslant 0$, we can show (see the appendix)

$$C_t^n \to C_t^\infty \equiv \int_0^t y_s e^{r(t-s)}\, ds \tag{27}$$

and

$$\frac{\mathscr{G}_t(\varepsilon_n, r_n, R_n)}{\varepsilon_n} \to \mathscr{G}_t^\infty \equiv 2U_t'(C_t^\infty) - (2K - r)\int_t^\infty U_s'(C_s^\infty) e^{-(\lambda - r)(s-t)}\, ds. \tag{28}$$

This indicates that, if the agent's optimal policy is indeed of the form of (22), then, for $\gamma > 0$, the time period, (t_n^*, t_n^{**}), of a mixed portfolio shrinks to a single point as $n \to \infty$, because $\mathscr{G}_t^\infty = 0$ holds for at most one $t \geqslant 0$. We thus expect that, for a sufficiently small ε and sufficiently large liquidity premium coefficient, the agent holds a mixed portfolio only for a short-time period, holds only the liquid consol before this time period, and holds only the illiquid consol after this time period. The following proposition confirms this intuition. Two of our results will use the following technical regularity condition, which ensures that an agent's current and

[19] If $y = 0$, an agent's control problem is stationary, and this continuous set of equations collapse into a single equation, which is why we can obtain an explicit analytic solution in Proposition 1.

future incomes are less important, relative to the wealth reserve, as his size of the wealth reserve increases.

Assumption 1. The income rate process y is strictly positive, differentiable, and satisfies

$$\frac{d}{dt}\left[\frac{y_t}{\int_0^t y_s e^{r(t-s)}\, ds}\right] < 0 \quad \text{for all } t. \tag{29}$$

Proposition 2. *Consider a sequence of economies defined by* $\{(\varepsilon_n, r_n, R_n)\}$, *with*

$$\varepsilon_n \to 0, \quad r_n \to r \quad \textit{and} \quad \frac{R_n - r_n}{2\varepsilon_n} \to K > 0 \quad \textit{as } n \to \infty. \tag{30}$$

(i) *If* $K < K_0 \equiv \beta + \lambda + \gamma r - r/2$, *then there exists some* N_1 *large enough that, for any* $n \geqslant N_1$, *an agent in the nth economy holds only the liquid consol throughout.*
(ii) *Under Assumption* 1, *if* $K = K_0$, *then for any time* $\bar{t} \in (0, \infty)$, *there exists some* $N_2 > 0$ *large enough that, in any economy* $n > N_2$, *an agent holds only the liquid consol for any time* $t < \bar{t}$.
(iii) *Under Assumption* 1, *if* $K > K_0$, *then there exists a time* $\tau \in (0, \infty)$ *such that, for any* $\delta \in (0, \tau)$, *there exists some* $N_3 > 0$ *large enough that in any economy* $n > N_3$, *an agent holds the liquid consol exclusively for any time* $t < \tau - \delta$, *and holds only the illiquid consol for any time* $t > \tau + \delta$. *Moreover,* τ *is given by*

$$\int_0^\infty \left(\frac{C^\infty_{\tau+t}}{C^\infty_\tau}\right)^{-\gamma} e^{-(\beta+\lambda-r)t}\, dt = \frac{2}{2K - r}, \tag{31}$$

where C^∞_t *is defined by* (27).

4.3. Equilibrium with small transactions costs

With the above solution of the optimal policy, we now consider the equilibrium formed by the continuum, $[0, 1]$, of agents in the limit of $\varepsilon \to 0$. An agent entering at time s has future income y_t at time $t + s$, where the process y is the same for all agents. We consider a sequence of economies defined by $\{\varepsilon_n\}$, and define an equilibrium in the nth economy as two rates of returns, r_n and R_n, such that the aggregate demand for each consol at any point of time is equal to the aggregate supply of that consol, in the sense of (19). We again conjecture that, for any n, in the equilibrium of the nth economy, the prices of the liquid and illiquid consols remain constant over time, and therefore that the rates of return of the two consols are constant over time.

We begin by constructing the equilibrium for the limit economy. Let r denote the equilibrium rate of return (in the limit economy) of the liquid consol and let K

denote the equilibrium liquidity-premium coefficient (in the limit economy). The total market wealth is equal to the total endowment in the limit economy, which implies, again using the law of large numbers, that

$$\frac{d}{r} = \int_0^\infty C_t^\infty e^{-\lambda t} \lambda \, dt = \frac{\lambda}{\lambda - r} \int_0^\infty y_t e^{-\lambda t} \, dt, \tag{32}$$

where we have again assumed that the age distribution of the population is at its steady state (exponential density) at time 0.

Based on Proposition 2, we conjecture that there exists some $\tau > 0$ such that agents in the limit economy hold only the liquid consol before "age" τ and hold only the illiquid consol after "age" τ. The fraction of the total market wealth held by agents with age less than τ is equal to the fractional supply of the liquid consol k. That is,

$$\int_0^\tau C_t^\infty e^{-\lambda t} \lambda \, dt = k \int_0^\infty C_t^\infty e^{-\lambda t} \lambda \, dt, \tag{33}$$

where C_t^∞ is given by (27). The equilibrium liquidity premium coefficient K is then related to τ by (31) in Proposition 2.

The above conjecture can be verified. We can show that the liquidity-premium coefficient for small ε is indeed close to the solution for K given by (33) and (31).

Proposition 3. *Suppose that* $W_0 = 0$ *and that Assumption* 1 *holds. Let* r_n *and* R_n *denote the equilibrium rates of return on the liquid and illiquid consols, respectively, in the nth economy with transactions cost coefficient* ε_n. *Then*

$$\lim_{n \to \infty} r_n = r \quad \textit{and} \quad \lim_{n \to \infty} \frac{R_n - r_n}{2\varepsilon_n} = K,$$

where r is given by (32), *and*

$$K = \frac{r}{2} + \left[\int_0^\infty \left(\frac{C_{\tau+t}^\infty}{C_\tau^\infty} \right)^{-\gamma} e^{-(\beta+\lambda-r)t} \, dt \right]^{-1} \tag{34}$$

with τ *given by* (33).

We again choose r exogenously, since it depends to a large extent on the exogenously chosen parameters d and y through (32).

4.4. Discussion and numerical calculation

With the above characterization of the equilibrium with small transactions costs, we now look at how the liquidity-premium coefficient depends on various economic factors.

First, the liquidity-premium coefficient increases with the agents' common risk-aversion coefficient γ. This is reasonable since part of the liquidity premium is the risk premium that agents require for holding the illiquid asset.

Secondly, the liquidity-premium coefficient increases as the relative supply, k, of the liquid consol decreases. This is again reasonable since, as the liquid consol supply decreases, the marginal holders of the illiquid consol have had less time to accumulate his wealth reserve and face a more uneven future income for liquidity shocks at different times, and thus demand a higher risk premium for holding the illiquid consol. This effect disappears if agents are risk-neutral. That is, if $\gamma = 0$, the liquidity-premium coefficient does not depend on the liquid consol supply.

Both of these effects can be illustrated by a numerical example with a constant income flow, $y_t = y_0$. For the limiting economy (with $\varepsilon \to 0$) under this specification, we have

$$C_t^{\infty} = \frac{y_0}{r}[e^{rt} - 1], \tag{35}$$

and the "age" of the marginal agent for the illiquid consol, τ, is determined, after simplifying (33), by

$$k = 1 + \frac{\lambda - r}{r - \delta} e^{-(\lambda-\delta)\tau} - \frac{\lambda - \delta}{r - \delta} e^{-(\lambda - r)\tau}. \tag{36}$$

The equilibrium liquidity-premium coefficient can then be calculated from (34) and (36). In Fig. 1, we present the numerical result for the following choice of parameters: $\lambda = 0.5$ (so that the expected arrival time of $1/\lambda$ matches the historical average holding time of securities traded on the New York Stock Exchange, which is about 2 years), $\beta = 0.05$, and $r = 5\%$. The liquidity-premium coefficient is calculated for all k values and for the relative risk aversion (RRA) coefficient γ in the range between 0 and 3.

As Fig. 1 shows, the liquidity premium depends strongly on the relative supplies of the liquid and illiquid securities and on the risk-aversion of investors. It increases as the supply of the liquid asset goes down and as agents' risk aversion goes up.

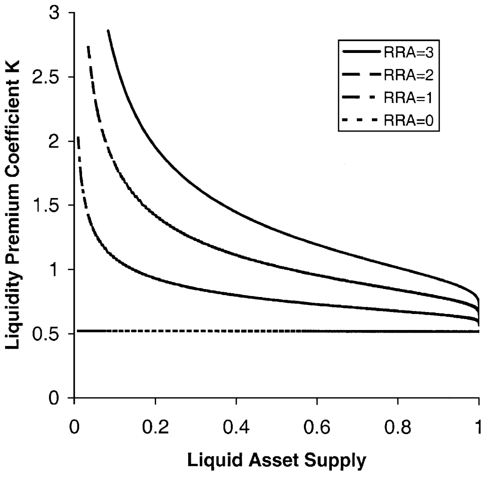

Fig. 1. The liquidity-premium coefficient, K, as a function of the total supply of the liquid asset. The four lines, from top to bottom, correspond to, respectively, different relative risk aversion coefficients (RRA) of $\gamma = 3$, $\gamma = 2$, $\gamma = 1$, and $\gamma = 0$. For other parameters, we have: $\lambda = 0.5$, $r = 0.05$, and $\beta = 0.05$.

Quantitatively, the liquidity-premium coefficient in our economy is much larger than that calculated under the PDV approach in which the price of the illiquid consol would simply be the price of the liquid consol adjusted for the present discounted value of all expected future transactions costs incurred by the marginal holder of the illiquid consol, with the riskfree rate as the discount rate (see, for example, [4]). If agents are risk-neutral, then our result is the same as that from the PDV approach, with a liquidity-premium coefficient of $\lambda + \beta - r/2 = 0.525$. In general, however, the PDV approach tends to underestimate the equilibrium liquidity premium when agents are risk-averse, face liquidity shocks, and are restricted from borrowing against future income.

We should also note that the liquidity premium obtained under the PDV approach is approximately the same as that obtained by assuming that agents have a fixed holding horizon of $1/\lambda$. The liquidity-premium coefficient generated by the latter approach is roughly λ, which is 0.5 in our numerical example.[20] So both of these approaches ignore the risk premium that agents demand for holding the illiquid asset, and therefore underestimate the required liquidity premium.

Our result can be helpful in understanding why market liquidity premium is quite high despite the fact that the market average holding horizon is relatively long. As discussed previously, given that the US stock market's average holding horizon is about 2 years, a simple PDV-based calculation would give $K \approx 0.5$, which is much smaller than those estimated by Amihud and Mendelson [4] and Brennan and Subrahmanyam [9] (between 1.5 and 2). Our result shows that investors, when facing liquidity shocks and being constrained from borrowing against future income, may indeed require a high liquidity premium. Of course, the highly stylized nature of this model does indeed limit the strength of this conclusion, but our paper nonetheless points to the importance of taking into account of liquidity shocks and borrowing constraints when we try to reconcile between the observed holding horizons and liquidity premia of securities in the market.

5. Conclusion

There is a gap between the empirical and theoretical literature on the effect of transactions costs on asset returns. While empirical studies have shown that transactions costs have a large effect on stock returns, theoretical studies generally predict only a small effect. This paper takes a small step in bridging such a gap by studying an equilibrium in which agents face surprise liquidity shocks and invest in liquid and illiquid riskless assets.

When investors face a random holding horizon due to liquidity shocks, the realized (net-of-transactions-cost) return of a riskless illiquid security can still be risky. In equilibrium, the required premium for bearing such risk depends on the

[20] For the technical setup in our model, the exact formula for the liquidity-premium coefficient demanded by an investor with a fixed horizon of T is shown by Vayanos and Vila [36] to be $K = 0.5r(1 + e^{-rT})/(1 - e^{-rT})$.

constraint that agents face when borrowing against future income; it is insignificant without borrowing constraint, but can be very high with borrowing constraint. Illiquidity, therefore, can have large effects on asset returns when investors face liquidity shocks and borrowing contraints. Specifically, equilibrium liquidity premium can be much higher than that demanded by investors with a fixed investment horizon that is equal to the expected arrival time of a liquidity shock. Our result can be helpful in understanding why some securities have high liquidity premia, despite their relatively low turnover frequency.

Acknowledgments

I thank Yakov Amihud, George Constantinides, Ken Singleton, Dimitri Vayanos, Jiang Wang, and, especially, Darrell Duffie, the associate editor, and two referees of this paper for many helpful comments. All errors are mine.

Appendix A. Proofs of lemmas and propositions

This appendix contains proofs for all the lemmas and propositions. We first define the set of feasible policies and then make some technical observations.

First, we note that it is never optimal for any agent to sell the illiquid consol prior to a liquidity shock since doing so amounts to paying the transactions costs earlier and earning less dividends. We therefore restrict out attention to policies with increasing A. A policy (a, A) is *feasible* if it satisfies (3)–(5) and A is non-decreasing. Secondly, from (3) to (5), it can be shown that all feasible policies satisfy the following *feasibility conditions*:

$$\delta_0 + (1+\varepsilon)\Delta_0 = 0, \tag{A.1}$$

$$d[e^{-rt}[\delta_t + (1+\varepsilon)\Delta_t]] = e^{-rt}[R - (1+\varepsilon)r]\Delta_t \, dt \tag{A.2}$$

and

$$\lim_{\substack{\varepsilon\to 0\\ R\to r}} C_t = \lim_{\substack{\varepsilon\to 0\\ R\to r}} (a_t + A_t) = C_t^\infty \equiv W_0 e^{rt} + \int_0^t y_s e^{r(t-s)}\, ds. \tag{A.3}$$

We next prove Lemma 1, which can be used to establish Proposition 1.

Proof of Lemma 1. The necessity of conditions (23)–(26) for the optimality of a policy (a, A) of the form of (22) was shown in their derivation. We now show that such a policy must be optimal by proving that no other feasible policy can produce a higher expected utility.

Let (a, A), with consumption $C = a + (1-\varepsilon)A$, denote a policy that is of the form of (22) and satisfies (23)–(26). Consider any other feasible policy

$(a+\delta, A+\Delta)$ with consumption $C+\xi = C+\delta+(1-\varepsilon)\Delta$. Using the concavity of U_t, we have

$$\int_0^\infty U_t(C_t+\xi_t)e^{-\lambda t}\lambda\, dt - \int_0^\infty U_t(C_t)e^{-\lambda t}\lambda\, dt$$
$$\leqslant \int_0^\infty U_t'(C_t)[\delta_t+(1-\varepsilon)\Delta_t]e^{-\lambda t}\lambda\, dt \tag{A.4}$$
$$= \mathcal{Q}_1+\mathcal{Q}_2+\mathcal{Q}_3+\mathcal{Q}_4, \tag{A.5}$$

where

$$\mathcal{Q}_1 \equiv \int_0^{t^*} U_t'(C_t)[\delta_t+(1+\varepsilon)\Delta_t]e^{-\lambda t}\lambda\, dt, \tag{A.6}$$

$$\mathcal{Q}_2 \equiv -\int_0^{t^*} U_t'(C_t)(2\varepsilon\Delta_t)e^{-\lambda t}\lambda\, dt, \tag{A.7}$$

$$\mathcal{Q}_3 \equiv \int_{t^*}^\infty U_t'(C_t)\left[\frac{1-\varepsilon}{1+\varepsilon}(\delta_t+(1+\varepsilon)\Delta_t)\right]e^{-\lambda t}\lambda\, dt, \tag{A.8}$$

$$\mathcal{Q}_4 \equiv \int_{t^*}^\infty U_t'(C_t)\left[\frac{2\varepsilon}{1+\varepsilon}\delta_t\right]e^{-\lambda t}\lambda\, dt. \tag{A.9}$$

Integrating (A.6) and (A.10) by parts and using the feasibility conditions for (a, A) and $(a+\delta, A+\Delta)$, we have

$$\mathcal{Q}_1 \equiv \int_0^{t^*}\left[\int_t^{t^*} U_s'(C_s)e^{-(\lambda-r)s}\lambda\, ds\right]e^{-rt}[R-(1+\varepsilon)r]\Delta_t\, dt, \tag{A.10}$$

$$\mathcal{Q}_3 == \mathcal{Q}_5 - \frac{1-\varepsilon}{1+\varepsilon}\int_{t^*}^\infty\left(\frac{R}{1+\varepsilon}-r\right)\left[\int_t^\infty U_s'(C_s)e^{-\lambda s}e^{\frac{R}{1+\varepsilon}(s-t)}\lambda\, ds\right]\delta_t\, dt, \tag{A.11}$$

where

$$\mathcal{Q}_5 \equiv \frac{1-\varepsilon}{1+\varepsilon}[\delta_{t^*}+(1+\varepsilon)\Delta_{t^*}]\int_{t^*}^\infty U_t'(C_t)e^{-\lambda t}e^{\frac{R}{1+\varepsilon}(t-t^*)}\lambda\, dt. \tag{A.12}$$

Combining (A.9) and (A.11),

$$\mathcal{Q}_3+\mathcal{Q}_4-\mathcal{Q}_5 = \frac{1}{1+\varepsilon}\int_{t^*}^\infty \mathcal{G}_t(\varepsilon, r, R)\delta_t e^{-\lambda t}\lambda\, dt, \tag{A.13}$$

with $\mathcal{G}_t$ defined in (8).

If $t^* = 0$, we have $\mathcal{Q}_5 = \mathcal{Q}_1+\mathcal{Q}_2 = 0$. If (a, A) satisfies (22)–(26), then $\mathcal{G}_t(\varepsilon, r, R) = 0$ for $t\in[0, t^{**})$, $\mathcal{G}_t(\varepsilon, r, R)\leqslant 0$ for $t\geqslant t^{**}$, and $\delta_t\geqslant 0$ for $t\geqslant t^{**}$. We thus have $\mathcal{G}_t(\varepsilon, r, R)\delta_t\leqslant 0$ for all t and $\mathcal{Q}_1+\mathcal{Q}_2+\mathcal{Q}_3+\mathcal{Q}_4\leqslant 0$. The lemma is then proved for $t^* = 0$.

If $t^* > 0$, applying feasibility conditions of (a, A) and $(a + \delta, A + \Delta)$ to (48), we have

$$\delta_{t^*} + (1+\varepsilon)\Delta_{t^*} = [R - (1+\varepsilon)r]e^{rt^*} \int_0^{t^*} \Delta_t e^{-rt}\, dt. \tag{A.14}$$

Substituting (A.14) into (A.12), then adding (A.7), (A.10), and (A.12) together, we have

$$\mathcal{Q}_1 + \mathcal{Q}_2 + \mathcal{Q}_5 = -\int_0^{t^*} \mathcal{L}_t\, \Delta_t e^{-\lambda t} \lambda\, dt, \tag{A.15}$$

where

$$\begin{aligned}\mathcal{L}_t \equiv\ & 2\varepsilon U_t'(C_t) - [R - (1+\varepsilon)r] \int_t^{t^*} U_s'(C_s) e^{-(\lambda - r)(s-t)}\, ds \\ & - \frac{1-\varepsilon}{1+\varepsilon}[R - (1+\varepsilon)r]e^{r(t^* - t)} \int_{t^*}^{\infty} U_s'(C_s) e^{\frac{R}{1+\varepsilon}(s-t^*)} e^{-\lambda(s-t)}\, ds.\end{aligned} \tag{A.16}$$

We define, for $t < t^*$,

$$\bar{\mathcal{K}}_t = -\int_t^{t^*} \mathcal{L}_s e^{-\lambda s} \lambda\, ds. \tag{A.17}$$

Integrating (A.15) by parts, we have

$$\mathcal{Q}_1 + \mathcal{Q}_2 + \mathcal{Q}_5 = \Delta_0 \bar{\mathcal{K}}_0 + \int_0^{t^*} \bar{\mathcal{K}}_t\, d\Delta_t. \tag{A.18}$$

Expanding (A.17) into three terms by using (A.16), exchanging the order of integration for the second term, integrating the third term, and comparing with (26), we have

$$\bar{\mathcal{K}}_t = e^{-\lambda t} \mathcal{K}_t. \tag{A.19}$$

Finally, substituting (A.19) into (A.18), and adding up (A.13) and (A.18),

$$\begin{aligned}&\mathcal{Q}_1 + \mathcal{Q}_2 + \mathcal{Q}_3 + \mathcal{Q}_4 \\ &\quad = \mathcal{K}_0 \Delta_0 + \int_0^{t^*} e^{-\lambda t} \mathcal{K}_t\, d\Delta_t + \frac{1}{1+\varepsilon} \int_{t^*}^{\infty} \mathcal{G}_t(\varepsilon, r, R)\, \delta_t e^{-\lambda t} \lambda\, dt.\end{aligned} \tag{A.20}$$

If (a, A) is of the form of (22) with $t^* > 0$, then $A_t = 0$ for $t < t^*$ and $a_t = 0$ for $t \geqslant t^{**}$. This implies that $\Delta_0 \geqslant 0$ and $\delta_t \geqslant 0$ for $t \geqslant t^{**}$. If, moreover, conditions (23)–(26) are met, then

$$\mathcal{Q}_1 + \mathcal{Q}_2 + \mathcal{Q}_3 + \mathcal{Q}_4 \leqslant 0. \tag{A.21}$$

The lemma is thus proved by (A.4) and (A.21). $\square$

Lemma 1 is used to show that the candidate policy in Proposition 1, of the form in (22), is indeed optimal.

Proof of Proposition 1. If $R \leqslant R^*$, the candidate policy is given by $A_t = 0$ and $C_t = a_t = W_0 \exp(rt)$, for all t. This policy is of the form of (22), with $t^* = t^{**} = \infty$, and for which we have

$$\mathscr{K}_t = \frac{\lambda W_0^{-\gamma} e^{-(\beta+\gamma r)t}(R - R^*)}{(\beta + \lambda + \gamma r - r)(\beta + \lambda + \gamma r)} \leqslant 0. \tag{A.22}$$

This policy therefore satisfies (23)–(26) and is optimal by Lemma 1.

If $R \geqslant R^{**}$, the candidate policy is given by (12) and (17), with $a_t = 0$ for all t. It is again of the form of (22), with $t^* = t^{**} = 0$, and, for all t,

$$\mathscr{G}_t(\varepsilon, r, R) = -U_t'(C_t)\frac{(1 + \varepsilon - 2\varepsilon\gamma)(R - R^{**})}{(1 + \varepsilon)(\beta + \lambda) + \gamma R - R} \leqslant 0. \tag{A.23}$$

This policy (a, A) therefore satisfies (23)–(26), and is optimal by Lemma 1.

If $R \in (R^*, R^{**})$, the candidate policy in (12) and (17) is of the form of (22), with $t^* = 0$ and $t^{**} = \infty$. From (8), we have $\mathscr{G}_t(\varepsilon, r, R) = 0$ for all t. So this policy satisfies (23)–(26), and is optimal by Lemma 1. $\square$

Proof of Proposition 2. If $K < K_0$, then there exists a N_1 such that, for any $n \geqslant N_1$, $R_n - r_n < 2K_0\varepsilon_n = R_n^* - r_n$, where R_n^* is defined as in (13) for the nth economy. Lemma 1 then implies that agents in the nth economy hold only the liquid consol.

If $K > K_0$. We first show that, for any $\delta > 0$, there exists N_3 such that, for any $n > N_3$, agents in the nth economy always find it optimal to hold only the illiquid consol for any time after $\tau + \delta$, regardless of their policy before time $\tau + \delta$. Let $(a_{(\tau+\delta)-}, A_{(\tau+\delta)-})$ denote an agent's (non-negative) holdings immediately before time $\tau + \delta$. Using a superscript n to denote the nth economy, our candidate policy (a^n, A^n) is

$$\begin{cases} a_t^n = 0, \\ A_t^n = [\frac{a^n_{(\tau+\delta)-}}{1+\varepsilon_n} + A^n_{(\tau+\delta)-}]e^{\frac{R_n}{1+\varepsilon_n}(t-\tau-\delta)} + \int_{\tau+\delta}^t \frac{y_s}{1 + \varepsilon_n} e^{\frac{R_n}{1+\varepsilon_n}(t-s)}\, ds, \end{cases} \quad t \geqslant \tau + \delta.$$

Consider any other feasible policy $(a^n + \delta^n, A^n + \Delta^n)$ for $t \geqslant \tau + \delta$. Using the concavity of U_t,

$$\begin{aligned} &\int_{\tau+\delta}^{\infty} U_t(C_t^n + \xi_t^n)e^{-\lambda t}\lambda\, dt - \int_{\tau+\delta}^{\infty} U_t(C_t^n)e^{-\lambda t}\lambda\, dt \\ &\quad \leqslant \int_{\tau+\delta}^{\infty} U_t'(C_t^n)[\delta_t^n + (1 - \varepsilon_n)\Delta_t^n]e^{-\lambda t}\lambda\, dt \\ &\quad = \mathscr{Q}_3^n + \mathscr{Q}_4^n, \end{aligned} \tag{A.24}$$

where

$$\mathscr{Q}_3^n = \int_{\tau+\delta}^{\infty} U_t'(C_t^n)\left[\frac{1 - \varepsilon_n}{1 + \varepsilon_n}(\delta_t^n + (1 + \varepsilon_n)\Delta_t^n)\right]e^{-\lambda t}\lambda\, dt, \tag{A.25}$$

$$\mathscr{Q}_4^n = \int_{\tau+\delta}^{\infty} U_t'(C_t^n)\left[\frac{2\varepsilon_n}{1 + \varepsilon_n}\delta_t^n\right]e^{-\lambda t}\lambda\, dt. \tag{A.26}$$

Applying feasibility conditions and integrating by parts, we have

$$\mathscr{Q}_3^n + \mathscr{Q}_4^n = \frac{1}{1+\varepsilon_n} \int_{\tau+\delta}^{\infty} \mathscr{G}_t(\varepsilon_n, r_n, R_n)\, \delta_t^n e^{-\lambda t} \lambda\, dt, \tag{A.27}$$

where

$$\mathscr{G}_t(\varepsilon_n, r_n, R_n) = U_t'(C_t^n)\Bigg[2\varepsilon_n - \frac{1-\varepsilon_n}{1+\varepsilon_n}[R_n - (1+\varepsilon_n) r_n] \times \int_t^{\infty} \frac{U_s'(C_s^n)}{U_t'(C_t^n)} e^{(\frac{R_n}{1+\varepsilon_n}-\lambda)(s-t)}\, ds\Bigg].$$

As $n \to \infty$,

$$\lim_{n\to\infty} \frac{\mathscr{G}_t(\varepsilon_n, r_n, R_n)}{2\varepsilon_n} = U_t'(C_t^{\infty})\left[1 - \frac{x-r}{2} \int_0^{\infty} \left(\frac{C_{t+s}^{\infty}}{C_t^{\infty}}\right)^{-\gamma} e^{-(\beta+\lambda-r)t}\, dt\right]. \tag{A.28}$$

Condition (29) implies that

$$\frac{\partial}{\partial t}\left(\frac{C_{t+s}^{\infty}}{C_t^{\infty}}\right) < 0, \tag{A.29}$$

which, combined with (31), implies that the limit in (A.28) is strictly negative for any $t > \tau$. This result, together with (A.24) and (A.27), proves our claim for the case of $K > K_0$ and $t \geqslant \tau + \delta$.

Continuing for the case of $K > K_0$, we show that for any $\delta \in (0, \tau)$, there exists N_3 such that, for any $n \geqslant N_3$, agents in the nth economy find it optimal to hold only the liquid consol for any time before $\tau - \delta$. We do so by proving that, for any given feasible policy $(\bar{a}^n, \bar{A}^n)$ which has $\bar{A}_t^n > 0$ for some $t < \tau - \delta$, the following revised policy, (a^n, A^n), is strictly utility enhancing:

$$\begin{cases} A_t^n = 0, \quad a_t^n = \int_0^t y_s e^{r_n(t-s)}\, ds & \text{if } t < \tau - \delta, \\ \dfrac{a_t^n}{a_t^n + A_t^n} = \dfrac{\bar{a}_t^n}{\bar{a}_t^n + \bar{A}_t^n} \equiv \theta_t^n & \text{if } t \geqslant \tau - \delta. \end{cases} \tag{A.30}$$

Let $\delta^n = \bar{a}^n - a^n$ and $\Delta^n = \bar{A}^n - A^n$. Using the concavity of U_t again,

$$\begin{aligned} &\int_0^{\infty} U_t(C_t^n + \xi_t^n) e^{-\lambda t} \lambda\, dt - \int_0^{\infty} U_t(C_t^n) e^{-\lambda t} \lambda\, dt \\ &\quad \leqslant \int_0^{\infty} U_t'(C_t^n)[\delta_t^n + (1-\varepsilon_n)\Delta_t^n] e^{-\lambda t} \lambda\, dt \\ &\quad = \tilde{\mathscr{Q}}_1^n + \tilde{\mathscr{Q}}_2^n + \tilde{\mathscr{Q}}_3^n, \end{aligned}$$

where

$$\tilde{\mathscr{Q}}_1^n = \int_0^{\tau-\delta} U_t'(C_t^n)[\delta_t^n + (1+\varepsilon_n)\Delta_t^n] e^{-\lambda t} \lambda\, dt, \tag{A.31}$$

$$\tilde{\mathscr{Q}}_2^n = \int_0^{\tau-\delta} U_t'(C_t^n)(2\varepsilon_n \Delta_t^n) e^{-\lambda t} \lambda\, dt, \tag{A.32}$$

$$\tilde{\mathscr{Q}}_3^n = \int_{\tau-\delta}^{\infty} U_t'(C_t^n)[\delta_t^n + (1-\varepsilon_n)\Delta_t^n]e^{-\lambda t}\lambda\, dt. \tag{A.33}$$

Integrating by parts and applying feasibility conditions, we have

$$\tilde{\mathscr{Q}}_1^n = \int_0^{\tau-\delta}\left[\int_t^{\tau-\delta} U_s'(C_s^n)e^{-(\lambda-r_n)s}\lambda\, ds\right]e^{-r_n t}[R_n-(1+\varepsilon_n)r_n]\Delta_t^n\, dt. \tag{A.34}$$

For the term $\tilde{\mathscr{Q}}_3^n$, from the feasibility conditions, we can show

$$\delta_t^n + (1-\varepsilon_n)\Delta_t^n = [\delta_{\tau-\delta}^n + (1+\varepsilon_n)\Delta_{\tau-\delta}^n]e^{r_n(t-\tau+\delta)} + o(\varepsilon_n), \quad t \geqslant \tau-\delta \tag{A.35}$$

and

$$\delta_{\tau-\delta}^n + (1-\varepsilon_n)\Delta_{\tau-\delta}^n = \delta_{(\tau-\delta)-}^n + (1+\varepsilon_n)\Delta_{(\tau-\delta)-}^n + o(\varepsilon_n). \tag{A.36}$$

Substituting (A.36) into (A.35) and then substituting (A.35) into (A.33),

$$\tilde{\mathscr{Q}}_3^n = [\delta_{(\tau-\delta)-}^n + (1+\varepsilon_n)\Delta_{(\tau-\delta)-}^n]\int_{\tau-\delta}^{\infty} U_t'(C_t^n)e^{r_n(t-\tau+\delta)}e^{-\lambda t}\lambda\, dt + o(\varepsilon_n). \tag{A.37}$$

Substituting (A.14), with $t^* = (\tau-\delta)-$, into (A.37), and then summing up (A.37) with (A.32) and (A.34),

$$\tilde{\mathscr{Q}}_1^n + \tilde{\mathscr{Q}}_2^n + \tilde{\mathscr{Q}}_3^n = -\int_0^{\tau-\delta} \Delta_t^n \mathscr{L}_t^n e^{-\lambda t}\lambda\, dt + o(\varepsilon_n), \tag{A.38}$$

where

$$\mathscr{L}_t^n = 2\varepsilon_n U_t'(C_t^n) - [R_n-(1+\varepsilon_n)r_n]\int_t^{\infty} U_s'(C_s^n)e^{-(\lambda-r_n)(s-t)}\, ds.$$

We have

$$\lim_{n\to\infty}\frac{\mathscr{L}_t^n}{2\varepsilon_n} = U_t'(C_t^{\infty})\left[1-\frac{2K-r}{2}\int_0^{\infty}\left(\frac{C_{t+s}^{\infty}}{C_t^{\infty}}\right)^{-\gamma} e^{-(\beta+\lambda-r)t}\, dt\right]. \tag{A.39}$$

Eqs. (A.29) and (31) imply that the limit in (A.39) is strictly positive for any $t<\tau$. This result, together with (A.30) and (A.38), proves our claim for the case of $K>K_0$ and $t<\tau-\delta$.[21]

Finally, for the case of $K=K_0$, the proof is identical to the proof for the case of $K>K_0$ and $t<\tau-\delta$, except with a change of $\tau-\delta$ into $\bar{t}$. $\square$

Proof of Proposition 3. In equilibrium, the total market wealth is equal to the total asset supply:

$$k\frac{d}{r_n} + (1-k)\frac{d}{R_n} = \int_0^{\infty}(a_t^n + A_t^n)e^{-\lambda t}\lambda\, dt.$$

Since investors hold both assets in equilibrium, we have $\lim_{n\to\infty}(R_n - r_n) = 0$. This, combined with (A.3), implies that $r = \lim_{n\to\infty} r_n$ is given by (32).

[21] Note that we have only shown what the optimal policy, if it exists, should be for $t<\tau-\delta$ and $t\geqslant\tau+\delta$. An optimal policy for $K>K_0$ can be shown to exist, at least among all policies that have a bounded rate of change for a and A.

Next we prove by contradiction that $\lim_{n\to\infty}(R_n - r_n)/\varepsilon_n = 2K$. Suppose that

$$X \equiv \limsup_{n\to\infty} \frac{R_n - r_n}{\varepsilon_n} \neq 2K. \tag{A.40}$$

Then there exists an increasing sequence $\{n_m : m \geqslant 1\}$ of integers such that

$$\lim_{m\to\infty} \frac{R_{n_m} - r_{n_m}}{\varepsilon_{n_m}} = X.$$

Let τ_X be defined by

$$\frac{X-r}{2} = \left[\int_0^\infty \left(\frac{C^\infty_{\tau_X+t}}{C^\infty_{\tau_X}}\right)^{-\gamma} e^{-(\beta+\lambda-r)t}\,dt\right]^{-1}.$$

Proposition 3 implies that, for any $\delta \in (0, \tau_X)$, there exists some $M > 0$ such that, for any $m > M$, agents in the n_mth economy holds only the liquid consol for $t < \tau_X - \delta$, and only the illiquid consol for $t \geqslant \tau_X + \delta$. We thus have

$$\frac{\int_0^{\tau_X-\delta} C_t^{(n_m)} e^{-\lambda t}\lambda\,dt}{\int_0^\infty (a_t^{(n_m)} + A_t^{(n_m)})e^{-\lambda t}\lambda\,dt} k \leqslant \frac{\int_0^{\tau_X+\delta} C_t^{(n_m)} e^{-\lambda t}\lambda\,dt}{\int_0^\infty (a_t^{(n_m)} + A_t^{(n_m)})e^{-\lambda t}\lambda\,dt}.$$

Taking the limit as $\delta \to 0$ and as $m \to \infty$,

$$k = \frac{\int_0^{\tau_X} C_t^\infty e^{-\lambda t}\lambda\,dt}{\int_0^\infty C_t^\infty e^{-\lambda t}\lambda\,dt} \tag{A.41}$$

which implies $\tau_X = \tau$ and thus $X = 2K$, a contradiction. So we have $\limsup_{n\to\infty}(R_n - r_n)/\varepsilon_n = 2K$. Similarly, $\liminf_{n\to\infty}(R_n - r_n)/\varepsilon_n = 2K$. We have thus finished the proof. $\square$

References

[1] A. Admati, P. Pfleiderer, A theory of intraday patterns: volume and price variability, Rev. Finan. Stud. 1 (1988) 3–40.

[2] S.R. Aiyagari, M. Gertler, Asset returns with transaction costs and uninsured individual risk: a stage III exercise, J. Monet. Econom. 27 (1991) 309–331.

[3] Y. Amihud, Illiquidity and stock returns, J. Finan. Markets 5 (2002) 31–56.

[4] Y. Amihud, H. Mendelson, Asset pricing and the bid-ask spread, J. Finan. Econom. 17 (1986) 223–249.

[5] Y. Amihud, H. Mendelson, Liquidity, maturity, and the yields on U.S. treasury securities, J. Finan. 46 (1991) 1411–1425.

[6] Y. Amihud, H. Mendelson, liquidity, asset prices and financial policy, Finan. Anal. J. (November–December 1991) 56–66.

[7] R.M. Anderson, Nonstandard methods with applications in economics, in: W. Hildenbrand, and H. Sonnenschein (Eds.), Handbook of Mathematical Economics, Vol. IV, Elsevier, North-Holland, Amsterdam, 1991.

[8] M.J. Brennan, T. Chordia, A. Subrahmanyam, Alternative factor specifications, security characteristics, and the cross-section of expected stock returns, J. Finan. Econom. 49 (1998) 345–373.

[9] M.J. Brennan, A. Subrahmanyam, Market microstructure and asset pricing: on the compensation for illiquidity in stock returns, J. Finan. Econom. 41 (1996) 441–464.

[10] T. Chordia, R. Roll, A. Subrahmanyam, Order imbalance, liquidity, and market returns, J. Finan. Econom., in press.
[11] T. Chordia, A. Subrahmanyam, V.R. Anshuman, Trading activity and expected stock returns, J. Finan. Econom. 59 (2001) 3–32.
[12] G.M. Constantinides, Capital market equilibrium with transaction costs, J. Polit. Econom. 94 (1986) 842–862.
[13] V.T. Datar, N.Y. Naik, R. Radcliffe, Liquidity and asset returns: an alternative test, J. Finan. Markets 1 (1998) 203–219.
[14] M.H.A. Davis, A.R. Norman, Portfolio selection with transactions costs, Math. Oper. Res. 15 (1990) 676–713.
[15] D. Duffie, T. Sun, Transaction costs and portfolio choice in a discrete-continuous time setting, J. Econom. Dyn. Control 14 (1990) 35–51.
[16] B. Dumas, E. Luciano, An exact solution to a dynamic portfolio choice problem under transactions costs, J. Finan. 46 (1991) 577–595.
[17] J. Dow, G. Gorton, Noise trading, delegated portfolio management, and economic welfare, J. Polit. Econom. 105 (1994) 1024–1050.
[18] A.L. Eisfeldt, Endogenous liquidity in asset markets, Working Paper, North-Western University, 2001.
[19] W.H. Fleming, S.G. Grossman, J.-L. Vila, T. Zariphopoulou, Optimal portfolio rebalancing with transactions costs, Working Paper, Brown University, 1990.
[20] E. Green, Individual-level randomness in a nonatomic population, Working Paper, University of Minnesota and Federal Reserve Bank of Minneapolis, 1994.
[21] S.J. Grossman, G. Laroque, Asset pricing and optimal portfolio choice in the presence of illiquid durable consumption goods, Econometrica 58 (1990) 25–52.
[22] J. Heaton, D. Lucas, Evaluating the effects of incomplete markets on risk sharing and asset pricing, J. Polit. Econom. 104 (1996) 443–487.
[23] C.M. Jones, A century of stock market liquidity and trading costs, Working Paper, Columbia University, 2001.
[24] A. Kyle, Continuous auctions and insider trading, Econometrica 53 (1985) 1315–1336.
[25] A. Lo, H. Mamaysky, J. Wang, Asset prices and trading volume under fixed transactions costs, Working Paper, MIT and Yale University, 2001.
[26] A. Lo, J. Wang, Trading volume: definitions, data analysis, and implications of portfolio theory, Rev. Finan. Stud. 13 (2000) 257–300.
[27] A.J. Morton, S.R. Pliska, Optimal portfolio management with fixed transaction costs, Math. Finan. 5 (1995) 337–356.
[28] T. Odean, Do investors trade too much?, Amer. Econom. Rev. 89 (1999) 1279–1298.
[29] Ľ. Pástor, R.F. Stambaugh, Liquidity risk and expected stock returns, J. Polit. Econom., in press.
[30] S.A. Ross, Intertemporal asset pricing, in: S. Bhattacharya, G.M. Constantinides (Eds.), Theory of Valuation: Frontiers of Modern Financial Theory, Rowman & Littlefield Publishers, Inc., Savage, MD, 1989, pp. 85–96.
[31] M. Schroeder, Optimal portfolio selection with fixed transaction costs, Working Paper, Michigan State University, 1998.
[32] S.E. Shreve, H.M. Soner, Optimal investment and consumption with transaction costs, Ann. Appl. Probab. 4 (1994) 609–692.
[33] W. Silber, Discounts on restricted stock: the impact of illiquidity on stock prices, Finan. Anal. J. (July–August 1991) 60–64.
[34] B. Tuckman, J.-L. Vila, Arbitrage with holding costs: a utility-based approach, J. Finan. 47 (1992) 1283–1302.
[35] D. Vayanos, Transaction costs and asset prices: a dynamic equilibrium model, Rev. Finan. Stud. 11 (1998) 1–58.
[36] D. Vayanos, J.-L. Vila, Equilibrium interest rate and liquidity premium with transaction costs, Econom. Theory 13 (1999) 509–539.