Available online at www.sciencedirect.com
SCIENCE DIRECT

Journal of Financial Economics 80 (2006) 531–562

JOURNAL OF Financial ECONOMICS
www.elsevier.com/locate/jfec

Dynamic liquidity in endowment economies ☆

Timothy C. Johnson*

London Business School, Regents Park, London, NWI 4SA, UK

Received 3 January 2005; received in revised form 14 March 2005; accepted 9 May 2005
Available online 4 January 2006

Abstract

This paper analyzes endogenous variations in aggregate liquidity that arise in standard representative-agent endowment economies. I introduce a natural definition of liquidity, essentially a shadow elasticity, that characterizes the price impact function or bid/ask spread that a small trader would experience. I compute this quantity for some tractable examples and uncover a rich variety of predictions that, in some cases, appear consistent with levels and covariations observed in the data. The results have important implications for the pricing and hedging of liquidity risk.

JEL classification: G12

Keywords: Liquidity; Liquidity risk; Asset pricing

1. Introduction

Liquidity risk in securities markets is a topic of growing concern to both investors and financial researchers. There is a widely shared sense that overall liquidity (however defined) does change over time, and that systematic drops, i.e., liquidity crises, are times of real economic stress. Such episodes tend to coincide with bad realizations of other important state variables: the market is likely to be down and volatility is likely to be up. However,

☆ My thanks to seminar participants at Boston College, Columbia University, Cornell, Dartmouth, London Business School, the Massachusetts Institute of Technology, the Norwegian School of Management, and the University of Toronto. I have also received very helpful comments and suggestions from Viral Acharya, Lasse Pedersen, and an anonymous referee.

*Fax.: +44 207 724 3317.
E-mail address: tjohnson@london.edu.

0304-405X/$ - see front matter © 2005 Published by Elsevier B.V.
doi:10.1016/j.jfineco.2005.05.004

the additional disutility that arises specifically from being unable to trade in these states (or from being able to do so only at very high cost) seems to present a distinct and worrisome dimension of risk.

This line of reasoning has led to recent investigation of the consequences of systematic fluctuations in liquidity (Pástor and Stambaugh, 2003; Acharya and Pedersen, 2004) for the pricing of assets. Yet viewing liquidity as an exogenous risk factor raises some important issues. Ultimately, the willingness of some agents to accommodate the trades of others must be an endogenous property of any economy; however, to the extent that liquidity is determined by the primary state variables that determine prices and discount rates, it does not represent a separate exposure to investors, and hence cannot be priced. That leaves other unmodelled factors—perhaps asymmetric information, search costs, or credit constraints—as possible sources of exogenous variation. That said, one cannot disentangle the role of exogenous factors without first understanding how endogenous factors affect liquidity determination.

Accordingly, this paper steps back from asset pricing to ask how marketwide liquidity ought to be related to fundamentals in the simplest, frictionless economies. The principle observation is that there is a direct measure, even in these settings, of the sensitivity of prices to trades. This measure is related to the slope of the representative agent's demand curve, and hence captures the price impact function, or market depth, faced by a small investor. Thus, this measure provides both a natural gauge of aggregate liquidity and an immediate quantification of how that liquidity is driven by the underlying sources of uncertainty.

Armed with this observation, we can then investigate the extent to which any given model can account for observed levels and dynamic properties of the liquidity of an endowment claim, for instance. This inquiry can be viewed in two ways. It can be taken as an assessment of these models themselves along a new empirical dimension, but it can also be interpreted as a quantification of the residual contribution of omitted factors to liquidity, and hence of the remaining exogenous liquidity risk implied by the model.

I illustrate the computation in several example economies, which are chosen mainly for expository purposes. It turns out that, even in these simple models, fundamental liquidity can resemble actual, measured liquidity in several ways. First, it may be of the right order of magnitude. Second, it may covary in the right degree with volatility and the level of the market. Third, and most important, in models with time-varying expected returns, illiquidity may covary positively with this expectation. Hence, even though there is no distinct liquidity risk in these economies, liquidity can appear to be a separate state variable and it can appear to be priced.[1]

The study of the determinants of liquidity at the aggregate level is still in its infancy. The basic fact that there is a systematic component to the liquidity of individual assets has been established by Chordia et al. (2000) and Huberman and Halka (2001), among others.[2] The authors of both of these papers state flatly that, to their knowledge, there were no extant theories of the sources of this comovement. While there is, of course, a huge microstructure

[1]In a similar critique, Novy-Marx (2004) suggests that omitting *any* risk factor from an asset pricing model will lead to the spurious finding of priced liquidity risk.

[2]Fleming (2001) documents commonality in the liquidity of US Treasury securities. Hasbrouck and Seppi (2001) find insignificant evidence of such comovement in a sample of 30 stocks in 1994. Chordia et al. (2004) document correlation between the systematic components of liquidity in stock and bond markets.

literature on liquidity provision, models based on inventory costs, search, or asymmetric information do not generally speak to the economywide determinants of their driving factors.[3] Yet empirical work using a variety of proxies has turned up a number of consistent, and intriguing, dynamic properties of aggregate liquidity.

Huberman and Halka (2001) and Jones (2002) document that, both at daily and yearly frequency, liquidity shocks are strongly positively correlated with market returns and that market returns predict future liquidity. Similarly, Huberman and Halka (2001) and Fujimoto (2004) find that market volatility is positively correlated with current and future illiquidity. Fujimoto (2004) also reports that illiquidity is higher when the Federal Funds rate is higher, monetary policy is tight, or the economy is in a recession. Finally, evidence of a correlation with expected returns appears in Jones (2002) and Amihud (2002), among others. These papers report significant, positive forecasting power for future market returns using proxies for current illiquidity.

A primary contribution of this paper is to propose a rationale for these findings that has nothing to do with microstructural affects or market imperfections. Remarkably, the facts above can all be understood purely through endogenous risk effects in frictionless representative agent models. In fact, the characterization of liquidity introduced here makes it clear that rather than a paucity, there is a plethora of theories: every standard equilibrium model imbeds predictions about the willingness of the market to accommodate trade. While the paper does not advance any one particular specification, it does suggest an approach towards building fuller descriptions by exploiting these inherent implications.

The rest of the paper is organized as follows. In the Section 2, I define the notion of liquidity that is studied here, and discuss its interpretation. While the definition is set in the context of an endowment economy without explicit trade, I show that it is consistent with corresponding economies in which the market provides liquidity to unmodelled agents who may wish to trade. Section 3 then computes the theoretical measure for three example models. I examine the resulting quantities analytically and numerically and argue that some parameterizations provide potential explanations for observed systematic liquidity patterns. Section 4 concludes by considering extensions of the work and summarizing the main implications.

2. A characterization of liquidity

This section introduces a new characterization of the liquidity of a securities market, which does not involve frictions, constraints, or asymmetric information. As such, it clearly ignores important aspects of liquidity determination. The claim advanced here is just that the proposed measure may serve as a useful starting point in understanding how liquidity is connected to economic fundamentals.

The first subsection provides the basic definition and discusses some choices of implementation. The following subsection considers the interpretation of the measure and shows how it can be related to the actual liquidity experienced by agents in the economy who need to trade.

[3]An important exception is Eisfeldt (2004) who offers a model of a market for claims to new projects whose quality varies over the business cycle, resulting in endogenous variation in information risk.

2.1. Definition

The setting the paper will consider is a frictionless pure exchange economy populated by a set of agents, $\mathscr{P}$, that can be characterized by a representative agent with standard preferences. There is a single perishable consumption good, the flow of which comes from $N+1$ productive assets indexed by $i \in \{0, 1, 2, \ldots, N\}$. There are $X^{(i)} > 0$ shares of the ith asset, each one receiving the payout $D_t^{(i)}$ per unit time. Total consumption flow is thus $C_t = \sum_i X^{(i)} D_t^{(i)}$.

In continuous time, the representative agent's lifetime expected utility is

$$J_t \equiv \mathrm{E}_t\left[\int_t^T \mathrm{e}^{-\phi(s-t)}\, u(C_s)\,\mathrm{d}s\right],$$

where ϕ is the subjective discount rate and $u()$ is any increasing, concave utility function and the time horizon may be infinite. The marginal utility process is denoted by

$$\Lambda_t \equiv \mathrm{e}^{-\phi t}\, u'(C_t),$$

and the standard result is that (under suitable regularity conditions)

$$P_t^{(i)} = \mathrm{E}_t\left[\int_t^T \Lambda_s\, D_s^{(i)}\,\mathrm{d}s\right] \Big/ \Lambda_t, \tag{1}$$

where $P^{(i)}$ is the agent's marginal valuation of shares of the ith asset, measured in consumption units:

$$P_t^{(i)} = \frac{1}{u'(C_t)} \frac{\partial J_t}{\partial X^{(i)}}. \tag{2}$$

Whether or not agents actually trade shares with one another $P^{(i)}$ may still be interpreted as a transactions price in the sense that any agent would be indifferent to a marginal change $\Delta X^{(i)}$ in his holdings of asset i if offset by a change $\Delta X^{(j)} = -\Delta X^{(i)} P^{(i)}/P^{(j)}$ in his holdings of any other asset j. Just as $P^{(i)}$ is the shadow price that would apply to a small, out-of-equilibrium perturbation to his portfolio, in the same way we can define the shadow elasticity as the *change* in $P^{(i)}$ that such a perturbation would induce.

To fix notation, I designate a single asset—asset zero—as the medium of exchange for the economy and focus on the liquidity of asset one. The consumption good may also intermediate trade, and this is perfectly compatible with the definitions below. The choice of the medium of exchange is not without loss of generality, however. Most naturally, one would think of it as "cash" or a money market account, i.e., something whose value is close to constant. Another possibility would be to view it as having the characteristics of "the market" so that the exchange models a trade-off of a specific security for generic wealth. Different choices will lead to differing measures of liquidity. Distinguishing between them is an empirical matter.

Now write $P^{(1)}(X^{(0)}, X^{(1)}) = P^{(1)}(X^{(0)}, X^{(1)}, \ldots, X^{(N)}; \mathscr{D})$, where $\mathscr{D}$ denotes all other state variables in the economy and the time subscript is suppressed for notational simplicity. The sensitivity that we want to quantify is the variation in $P^{(1)}$ when *both* $X^{(1)}$ and $X^{(0)}$ are changed so as to leave the agent indifferent, since these are the marginal variations that

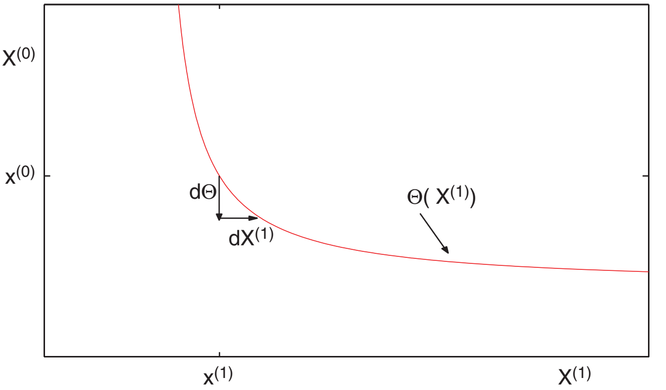

Fig. 1. The figure plots a particular level surface for the value function given by the two-period example in Section 2.1 whose arguments are the quantities, $X^{(0)}$ and $X^{(1)}$, of the two assets. The isoquant implicitly defines the function $\Theta(x)$ and its derivative which appear in the definition of illiquidity.

define the price system.[4] To define illiquidity we need to describe this perturbation in terms of how $X^{(0)}$ changes when we change $X^{(1)}$ by $\mathrm{d}X^{(1)}$.

Intuitively, value-neutral exchanges are simply ones that leave J, the agent's maximized lifetime objective—viewed as a function of the endowments—unchanged. Thus, the variations in the share holdings must be along the level curves of $J(X^{(0)}, X^{(1)})$ in the $(X^{(0)}, X^{(1)})$ plane; see Fig. 1. Mathematically, this curve can be parameterized by the implicit function $X^{(0)} = \Theta(X^{(1)})$ defined by the restriction that $J(\Theta(X^{(1)}), X^{(1)})$ equals the value obtained at the actual endowment (called $(x^{(0)}, x^{(1)})$ in the graph). Equivalently, If the mappings involved are all smooth, then Θ can be viewed as being defined as the solution to the equation $\mathrm{d}J(\Theta(X^{(1)}), X^{(1)})/\mathrm{d}X^{(1)} = 0$ subject to $x^{(0)} = \Theta(x^{(1)})$.

We can now see precisely how the restriction to the indifference curve Θ constrains $\mathrm{d}X^{(0)}/\mathrm{d}X^{(1)}$ or $\mathrm{d}\Theta(X^{(1)})/\mathrm{d}X^{(1)}$. Because

$$\frac{\mathrm{d}J(\Theta(X^{(1)}), X^{(1)})}{\mathrm{d}X^{(1)}} = 0 \Rightarrow \frac{\partial J(\Theta(X^{(1)}), X^{(1)})}{\partial X^{(1)}} = -\frac{\partial J(\Theta(X^{(1)}), X^{(1)})}{\partial X^{(0)}} \frac{\mathrm{d}\Theta(X^{(1)})}{\mathrm{d}X^{(1)}},$$

substituting from Eq. (2) we must have

$$\frac{\mathrm{d}\Theta(X^{(1)})}{\mathrm{d}X^{(1)}} = \frac{\mathrm{d}X^{(0)}}{\mathrm{d}X^{(1)}} = -\frac{P^{(1)})}{P^{(0)}}. \tag{3}$$

Thus, a value-neutral exchange is just one in which the relative quantities change in inverse proportion to their relative prices.[5] With this observation, we can readily specify the elasticity, or price impact, that characterizes such a trade.

Definition 2.1. *The illiquidity* $\mathscr{I} = \mathscr{I}^{(1,0)}$ *of asset one with respect to asset zero is the elasticity*

$$\mathscr{I} = -\frac{X^{(1)}}{P^{(1)}} \frac{\mathrm{d}P^{(1)}(\Theta(X^{(1)}), X^{(1)})}{\mathrm{d}X^{(1)}} = -\frac{X^{(1)}}{P^{(1)}} \left(\frac{\partial P^{(1)}}{\partial X^{(1)}} - \frac{P^{(1)}}{P^{(0)}} \frac{\partial P^{(1)}}{\partial X^{(0)}} \right), \tag{4}$$

[4]The indifference condition is also equivalent to restricting attention to perfectly competitive exchanges in a setting that supports actual trade. This is the approach taken in the disaggregation results in the next subsection.

[5]Note that the exchange need not be *wealth* neutral in that, although the trade is "self-financing" $(P^{(0)}\,\mathrm{d}X^{(0)} = -P^{(1)}\,\mathrm{d}X^{(1)})$, this does not entail $\mathrm{d}P^{(0)}\,X^{(0)} = -\mathrm{d}P^{(1)}\,X^{(1)}$. The capital gains terms need not offset.

where $(\Theta(x), x)$ is the locus of endowment pairs satisfying $J(\Theta(x), x) = J(X^{(0)}, X^{(1)})$. Liquidity, $\mathscr{L}$, may be defined as $1/\mathscr{I}$.

The second equality in the definition above follows from Eq. (3) and makes clear that one need not compute the function Θ in order to compute $\mathscr{L}$ or $\mathscr{I}$.[6] However, using Θ, one could also derive the price impact resulting from any noninfinitesimal trade that is subject to the value-neutral condition. The inverse of the function $P^{(1)}(\Theta(X^{(1)}), X^{(1)})$ also has the interpretation as the representative agent's demand curve for asset one, subject to the value-neutral constraint. If this demand curve is flat, liquidity is infinite under the above definition, as it should be. But this need not hold. In general, changes in the representative agent's wealth composition will alter marginal valuations via changes to the dynamics of aggregate consumption.

As a simple illustration, consider an economy with two securities, each of which pays a single lump-sum dividend. Call the payment dates T_0 and T_1 and denote the securities asset zero and asset one. If we are at t and current consumption C_t is unaffected by the portfolio holdings, then $\mathrm{d}P^{(1)}/\mathrm{d}X^{(1)}$ is

$$\frac{\mathrm{d}}{\mathrm{d}X^{(1)}}\mathrm{E}\left(\mathrm{e}^{-\phi T_1}\frac{u'(X^{(1)}D^{(1)})}{u'(C_t)}D^{(1)}\right) = \mathrm{E}\left(\mathrm{e}^{-\phi T_1}\frac{u''(X^{(1)}D^{(1)})}{u'(C_t)}(D^{(1)})^2\right).$$

With power utility, this is just $-\gamma P^{(1)}/X^{(1)}$, in which case $\mathscr{I} = \gamma$.[7]

In this example, liquidity just equals the elasticity of intertemporal substitution, $\mathscr{L} = 1/\gamma$. This is intuitive: a high elasticity of intertemporal substitution implies that the representative agent will readily transfer consumption from T_0 to T_1. Hence, a request for such a transaction from a small investor has only a limited impact on prices, i.e., liquidity is high. A low elasticity of intertemporal substitution implies that the representative agent is not willing to accommodate such intertemporal shifts in consumption, and thus $\mathscr{L}$ is low. Interestingly, in the power case, liquidity does not depend at all on the risk characteristics of the two securities or the timing of their payouts.

When the quantities of both assets enter each period's utility—as is the case in all the models in Section 3—more interactions come into play. Suppose, in the above example, that $T_0 < T_1$ and T_0 consumption affects T_1 marginal utility via a habit-type term $u'(C_1) = (X^{(1)}D^{(1)}/X^{(0)}D^{(0)})^{-\gamma}$. Then $\mathrm{d}P^{(1)}/\mathrm{d}X^{(1)}$ is

$$\mathrm{E}\left(\mathrm{e}^{-\phi T_1}\frac{u''(C_1)}{u'(C_t)}D^{(1)}\frac{\mathrm{d}R}{\mathrm{d}X^{(1)}}\right) = -\gamma\mathrm{E}\left(\mathrm{e}^{-\phi T_1}\frac{u'(C_1)}{u'(C_t)}\frac{D^{(1)}}{R}\frac{\mathrm{d}R}{\mathrm{d}X^{(1)}}\right),$$

where $R = X^{(1)}D^{(1)}/X^{(0)}D^{(0)}$. Since

$$\frac{1}{R}\frac{\mathrm{d}R}{\mathrm{d}X^{(1)}} = \left(\frac{1}{X^{(1)}} - \frac{1}{X^{(0)}}\frac{\mathrm{d}X^{(0)}}{\mathrm{d}X^{(1)}}\right)$$

[6]The definition does assume that the derivatives in the elasticity exist. Note that the elasticity may still be well defined (in the limit) for an asset of net zero supply.

[7]Although it is not needed in the computation, the level curve Θ of the value function can be explicitly derived since $J(X^{(0)}, X^{(1)})$ is just the sum of separate terms in $X^{(0)}$ and $X^{(1)}$. The curve in Fig. 1 is an example of the solution with $\gamma = 2$.

from the value-neutrality condition, we therefore have

$$\frac{\mathrm{d}P^{(1)}}{\mathrm{d}X^{(1)}} = -\gamma \frac{P^{(1)}}{X^{(1)}} \left(1 + \frac{P^{(1)}X^{(1)}}{P^{(0)}X^{(0)}}\right) \Rightarrow \mathscr{I} = \gamma \left(1 + \frac{P^{(1)}X^{(1)}}{P^{(0)}X^{(0)}}\right).$$

Here, the intertemporal substitution effect is magnified by a factor that depends on the relative weights of the two assets in total wealth. The more valuable time-T_1 consumption is, the more the agent is disinclined to substitute away from it. Paying away time-T_0 wealth decreases T_1 marginal utility via lowering the anticipated habit level.

These simple examples show that, in general, the elasticity of substitution between two assets is a nontrivial quantity that may depend on the state of the economy. The second version also illustrates that the dynamic properties of *both* assets may affect $\mathscr{I}$. The same security may be more or less liquid depending on what it is exchanged for.

It should also be noted that the choice of the medium of exchange is distinct from the choice of numeraire. The definition of $\mathscr{I}$ uses the pricing function denominated in units of consumption, as usual. However, if trades are actually conducted using asset zero, another possibility is to measure prices in units of that asset.

Definition 2.2. *The nominal illiquidity* $\bar{\mathscr{I}}^{(1,0)}$ *of asset one with respect to asset zero is the illiquidity (as defined above) of the nominal price* $\bar{P}^{(1)} \equiv P^{(1)}/P^{(0)}$.

Again, the choice of real versus nominal measures is not without loss of generality. For empirical work, the nominal definition may be more appropriate. Of course, the two definitions coincide when asset zero is riskless in the sense that its real price is constant. In Section 3, I compute real rather than nominal liquidities because it is standard practice to measure prices in consumption units. Also, in examining covariations with other dynamic quantities (such as volatilities and expected returns), I do not want to have to disentangle real and nominal effects. To the extent that asset zero in the examples is unvolatile (hence ‘‘money-like’’), the difference will be small.

Before turning to computations, it is worthwhile to consider some issues of interpretation to understand how $\mathscr{I}$ relates to the actual transactions costs experienced by market participants.

2.2. Interpretation

Why is the slope of the representative agent's demand curve a measure of liquidity? The intuition is simple: a small investor who needs to trade $\Delta X^{(1)}$ for whatever reason faces a market supply curve given by the total supply minus the representative agent's demand. Thus, the (inverse) slope of this curve answers the question: how much will the price move against him for each share he trades? Alternatively, if the representative agent were to quote a two-sided market for a fixed number of shares, the investor would face a spread proportional to $\mathscr{I}$.

There is a long tradition in the microstructure literature, dating from Garman (1976) and Stoll (1978), that similarly views liquidity as being provided by a risk-averse market maker who is compensated for the ‘‘inventory cost’’ of holding a suboptimal portfolio. So the identification is not new. However, the classical formulations are partial equilibrium ones in which the exogenous institutional arrangements (which determine who may be a market maker and how they can compete and hedge) induce a cost structure that is essentially overlayed on an underlying fundamental price. The observation here is that there is a limiting characterization of the inventory cost that is a property of the equilibrium itself, and that is independent of the microstructural details.

In effect, the interpretation in this paper is equivalent to viewing the market maker as the representative agent. The disaggregated picture of liquidity evolution, then, would involve subsets of agents in the overall population experiencing needs to trade at different times. The rest of the population accommodates them, and share holdings are redistributed accordingly. If the liquidity demanders are collectively small, then we can, to some extent, regard this whole process as going on in the background of any representative agent economy.

This story can be formalized in different ways. While I make no attempt to treat the topic comprehensively, several issues arise that merit discussion.

First, what happens if liquidity demands are nonvanishing in aggregate? To be sure, we can regard the liquidity trade as occurring "in the background" when each trade is infinitesimal and the total mass of all such trades at each time approaches zero. But it is also interesting to ask how a given representative agent model would evolve (and how the computation of $\mathscr{I}$ might change) when liquidity needs are finite, and perhaps even large, as they may be in practice.

A partial answer is given by the following proposition, which envisions discrete trade at a sequence of dates $\{\tau_k\}_{k=1}^{K}$ between subsets of agents.

Proposition 2.1. *Suppose at each trade date a subset $\widetilde{\mathscr{P}}_k$ of agents is required to change their holdings of asset one by $\pm\varepsilon$ in exchange for asset zero (the sign of the trade to be revealed at τ_k) and to hold the resulting portfolio $\widetilde{X}_{\tau_k}$ until τ_{k+1}.*

Assume the remaining agents, $\widehat{\mathscr{P}}_k$, are required to competitively establish bid and ask prices $P_k^{(1,B)}$ and $P_k^{(1,A)}$ at which the trades can occur. Further assume that the preferences and beliefs of each $\widehat{\mathscr{P}}_k$ can be by characterized by those of a common representative agent $\widehat{\mathscr{R}}$ whose utility function is $u(\,)$.

Denote the holdings of $\widehat{\mathscr{P}}_k$ by $\widehat{X}_t = X - \widetilde{X}_\tau$, where $\tau = \max\{\tau_k : \tau_k \leqslant t\}$. Let $\widehat{C}_{t,s} = \sum_{i=0}^{N} \widehat{X}_t^{(i)} D_s^{(i)}$.

Then, at every date t, the value function of the representative agent is

$$J = \mathrm{E}_t\left[\int_t^T \mathrm{e}^{-\phi(s-t)}\, u(\widehat{C}_{t,s})\, \mathrm{d}s | \mathscr{D}_t\right] \equiv J(\widehat{X}; t, \mathscr{D}_t)$$

and the price $P_t^{(i)}$ of asset i is $1/u(C)\,\partial J/\partial X^{(i)}$ evaluated at $\widehat{C}_{t,t}$ and $\widehat{X}_t$.

Moreover, the bid–ask spread $Z \equiv P_k^{(1,A)} - P_k^{(1,B)}$ satisfies

$$\lim_{\varepsilon\to 0} \frac{\widehat{X}^{(1)}}{2\varepsilon}\frac{Z}{P^{(1)}} = \mathscr{I}, \tag{5}$$

where $\mathscr{I}$ is defined by Eq. (4).

Note: all proofs appear in the Appendix.

The proposition says that, if the subset of the population that provides liquidity stays "representative" over time, then we can view the economy as a standard, fixed-endowment one at each point, and we can compute prices (and liquidity) without regard to future portfolio flows resulting from liquidity demand. The only dynamic effects of the exchanges arise through the current position vector $\widehat{X}_t$. So, to the extent that the net effect of the trading keeps the portfolio shares of the liquidity providers stable through time, the economy will evolve like its fixed-share equivalent. Any perturbations in equilibrium quantities can be bounded by placing bounds on the deviation of this vector from its mean.

The proposition also exhibits (in Eq. (5)) precisely how $\mathscr{I}$ is related to the competitive bid–ask spread. Interestingly, the spread is not affected by any of the stochastic details of the trade process.[8] In fact, the proposition does not place conditions on when or why the liquidity shocks take place. One could imagine a variety of mechanisms underlying them. (For example, it could be random shocks to risk aversion, as in Campbell et al. (1993).) Here, the reasons underlying liquidity demand do not matter for determining elasticities of asset substitution. The economics of liquidity are all driven by the supply side.

In the story so far, the view of liquidity shocks is that of an enforced rearrangement of asset holdings (i.e., in the secondary market). A related question is whether the notion of liquidity quantified by $\mathscr{I}$ is also consistent with a world in which the supplies of the underlying assets actually do change. This question is addressed in the next proposition, which also explicitly describes a disaggregated exchange mechanism that implements the assumption of competitive liquidity provision that was used above.

Proposition 2.2. *Assume that the population $\mathscr{P}$ consists of M identical agents each having the initial endowments $\{x_0\} \equiv \{X_0/M\}$. Suppose that at each time t this population is required to participate in open* (*English*) *auctions for $+\varepsilon$ and $-\varepsilon$ shares of asset one* (*to be paid in shares of asset zero*), *where actual supply shocks will be realized at a sequence of trade dates $\{\tau_k\}_{k=1}^{K}$. At τ_k the sign of the shock is revealed, and every winning bid or offer is filled. Then there is a unique dominant strategy equilibrium with the following properties.*

(i) *At each date, every agent matches the best bid and ask prices, denoted $P_t^{(1,B)}$ and $P_t^{(1,A)}$, respectively, and trades the same number of shares.*
(ii) *If the value function is homogeneous of degree one in the endowments, the equilibrium price and limiting bid–ask spread are as given in Proposition* 2.1.

Proposition 2.2 again illustrates a justification for viewing $\mathscr{I}$ as the appropriate liquidity metric in a representative agent economy. In determining the value function (and hence prices and liquidity), the fact that the shares of the assets will be altered in the future does not matter prior to the actual exchange. All that agents need to know is that these supply shocks will be absorbed at perfectly competitive prices. Intuitively, agents are indifferent to any sequence of future trades each of which will leave them indifferent.

Like the first proposition, this one places no conditions on the process giving rise to the supply shocks. These shocks could be the result of irrational noise trading, or they could come from open market operations of a central bank adjusting the supply of the reference asset (i.e., money). They could also capture the primary issuance and repurchase of risky assets by the corporate sector. This observation implies that any simple endowment model (such as the ones examined below) can have an almost arbitrary supply process literally appended to it, thereby generating a family of alternative models. In each of these, the computation of prices and liquidity is the same, but is evaluated at the current supply vector X_t.

[8]There is, however, an implicit assumption that the representative agent's information set is not affected by the occurrence of the trade.

Juxtaposing the two propositions, we can conclude that the notion of liquidity studied here is the same whether one regards the supply shocks as coming from outside the economy or as arising within it. In either case, the measure $\mathscr{I}$ summarizes the willingness of the market to accommodate somebody else's need to trade. This capacity is the same for both primary and secondary markets.

It should be emphasized that the theory developed in this section sheds no light at all on *why* the need to trade arises, and hence says nothing about the determination of volume. Moreover, the demand for liquidity may, itself, be influenced by the cost of trading. It is also worth remembering that all other determinants of liquidity supply—such as credit conditions and asymmetric information—have not been taken into account.

This is certainly an oversimplified view of liquidity. The question now is: is it a useful one? In a sense, the contribution of this section is to point out that this is a question that can be legitimately posed to different asset pricing models to determine the extent to which their implied liquidity processes capture any of the salient features of the data. In the next section, I compute $\mathscr{I}$ in some example models and explore its analytical and numerical properties. While all of the parameterizations have shortcomings, there still appear to be important respects in which they may provide potential explanations for observed liquidity dynamics.

3. Some examples

To explore the systematic behavior of liquidity, as defined above, this section examines some tractable examples. The immediate goals are expository, namely, to illustrate the computation and to gain some economic intuition about what drives the changing willingness of agents to accommodate trades within these frictionless settings. The broader purpose of the exercise is to underscore the rich variety of predictions about liquidity dynamics embedded in apparently simple models.

Note that none of the examples here was intended to be a model of liquidity, and I do not attempt to formally test them.[9] Instead, I choose empirically motivated fundamental parameters to assess whether the economic mechanisms at work provide plausible accounts of actual liquidity patterns. I do intend to suggest that we can and should regard these models as offering falsifiable explanations for an important dynamic quantity. In this regard, it is encouraging that some of the formulations do, indeed, get a lot of the qualitative features right. The approach introduced here thus points the way towards the development of a fuller understanding of aggregate liquidity.

The examples build on recent work in asset pricing that explicitly models changes in the composition of the aggregate consumption stream. With multiple assets, or sectors, the relative contribution of each dividend stream to the aggregate will fluctuate with their relative growth rates. As noted in Section 2.1, liquidity effects arise in equilibrium precisely because marginal changes in the representative agent's portfolio also alter the dynamics of aggregate consumption in much the same way. Since market clearing imposes $C_t = \sum_i X^{(i)} D_t^{(i)}$, a perturbation to $X^{(i)}$ has a similar effect on $\mathrm{d}C_t$ to a stochastic change in $D^{(i)}$. Altering the dynamics of $\mathrm{d}C_t$, then, affects the stochastic discount factor in all future periods, thereby changing current valuations. Analyzing liquidity requires tracing

[9]Indeed, each of them can be rejected by standard return-based asset pricing tests.

through these general equilibrium effects.[10] The models below provide a natural and transparent framework for doing so.

All the examples are of two-asset economies with power utility. The subjective discount rate and the exponent of the utility function are called ϕ and γ, respectively. As in the previous section, I let $X^{(1)}$ and $D^{(1)}$ denote the number of shares and dividends-per-share of the asset of primary interest, with $X^{(0)}$ and $D^{(0)}$ denoting the same quantities for the asset that serves as the medium of exchange. The consumption share of the first asset is then $D^{(1)}X^{(1)}/(D^{(0)}X^{(0)} + D^{(1)}X^{(1)})$, which I call s. This ratio serves as a sufficient state variable for pricing, incorporating all necessary information about the relative supplies of the assets.

3.1. The model of Santos and Veronesi (2005)

Santos and Veronesi (2005) consider an economy in which the consumption share of each of $N+1$ assets is stationary. Their primary focus is the case in which one of the assets (analogous to my asset zero) is interpreted as human capital, so that its "dividend" represents labor income. For our purposes, asset zero should be interpreted as nominal money or perhaps a government bond, i.e., a security with low (but not zero) risk that intermediates exchanges.

To start, I focus on the case in which $N = 1$ and interpret asset one as "the market," i.e., a claim to all other wealth. In this case, there are only two securities, and the stochastic specification treats them essentially symmetrically. The total consumption process, $C_t = D_t^{(0)}X^{(0)} + D_t^{(1)}X^{(1)}$, obeys

$$\frac{\mathrm{d}C}{C} = \mu_C(s_t)\,\mathrm{d}t + \sigma_C\,\mathrm{d}W_t^C, \tag{6}$$

where σ_C is a constant, and the consumption share obeys

$$\mathrm{d}s = a(\bar{s} - s_t)\,\mathrm{d}t + bs_t(1 - s_t)\,\mathrm{d}W_t^s. \tag{7}$$

The parameters a, b, and $\bar{s}$ are positive constants, and $\bar{s} \in (0, 1)$ is the steady-state share of asset one. The only asymmetry between the assets arises from the correlation, ρ, between the Brownian motions $\mathrm{d}W_t^s$ and $\mathrm{d}W_t^C$. When $\rho > 0$ (as it is in all the cases below), the dividend stream of asset one covaries more with consumption than asset zero, making investors more concerned with its dividend risk. This is the sense in which asset zero can be interpreted as safer.

Santos and Veronesi (2005) show that if either the relative risk aversion parameter, γ, is one (i.e., log utility) or $\mu_C = \rho b \sigma_C s_t + \bar{\mu}$ ($\bar{\mu}$ being an arbitrary constant), then the value of each asset is linear in the total dividends of each. Specifically,

$$P_t^{(0)}\; X^{(0)} = B_{(0,0)} D_t^{(0)} X^{(0)} + B_{(0,1)} D_t^{(1)} X^{(1)}, \tag{8}$$

$$P_t^{(1)}\; X^{(1)} = B_{(1,0)} D_t^{(0)} X^{(0)} + B_{(1,1)} D_t^{(1)} X^{(1)}, \tag{9}$$

where the matrix elements $B_{(i,j)}$ are constant. Notice from Eq. (9) that the price per share of asset one has a term that depends on the dividend of the *other* asset, even though holders of

[10] As the referee points out, if consumption—and hence the stochastic discount factor—were taken as exogenous, it would appear that all assets were perfectly liquid. The shares $X^{(i)}$ only enter the pricing (1) through Λ.

asset one do not receive any such cash flow. The reason is that, since the two dividend streams are cointegrated, the level of $D^{(0)}$ today provides information about future values of $D^{(1)}$. In particular, the expected discounted value of the latter stream will be higher the higher $D^{(0)}$ because $D^{(1)}$ is drifting towards $\bar{s}(D^{(0)}(X^{(0)}/X^{(1)})+D^{(1)})$.

Given the simple linear structure of prices, $\mathscr{I}$ can be evaluated by straightforward differentiation (shown in the proof of the next proposition). The result is a ratio of quadratics in s that is not particularly revealing. However, some analytical results are immediate.

Proposition 3.1. *Assume the conditions of Proposition* 1 *in* Santos and Veronesi (2005) *and that* $L(s)=\phi-(1-\gamma)\mu_C(s)+\gamma(1-\gamma)\sigma_C^2/2>0$. *Then* $P^{(0)}$, $P^{(1)}$, *and all the elements of the matrix B are positive, and*

(i) $\mathscr{I}(s)>0, \quad \forall s<1$,
(ii) $\mathscr{I}(1)=0$, *and*
(iii) $\mathscr{I}(0)=(L(1)+a)/(L(1)+a(1-\bar{s}))>1$.

The first point of Proposition 3.1(i), shows that the risky asset does not have infinite liquidity whenever the ratio of its dividends to consumption is less than one. The intuition is as follows. Consider first the case in which this ratio is small ($s\approx 0$). Then the marginal value of this asset is high since its dividends are expected to grow at a faster rate than the economy. Since this growth rate decreases with s, prices will fall if units of the asset are added: it is illiquid. As the share s becomes larger, the same logic applies, only now the future growth rate accounts for a smaller share of the asset's valuation (current dividends account for more), so the sensitivity of the price to s declines. Indeed, the asset becomes perfectly liquid when $s=1$, as (ii) shows.

To illustrate the effects that arise in this model, I compute the illiquidity under different configurations that share the same, standard choices for the consumption and preference parameters, specifically, $\phi=0.01, \gamma=2, \sigma_C=0.03$, and $\bar{\mu}=0.03$. I also fix the consumption-share diffusion parameter at $b=0.6$ and its correlation with consumption at $\rho=0.2$. These choices make asset one significantly more risky than asset zero (in the sense discussed above), while keeping changes in the share small over short horizons.[11] In the base case, I also set $a=0.2$ and $\bar{s}=0.8$, in which case asset one accounts for 80% of consumption on average, with a rate of mean-reversion of about four years. If asset zero represents government bonds, this describes an economy whose debt share is similar to that of the US and that varies at business cycle frequencies.

To start, however, I consider the effect of varying the last two parameters. Fig. 2 shows $\mathscr{I}$ as a function of s for different values of $\bar{s}$.

The plot yields several observations. The most basic is that illiquidity can be economically significant. For each of the parameterizations shown, $\mathscr{I}\approx 1$ when $s=\bar{s}$. In other words, a trade in 1% of an asset's supply would move prices by 1%. This is too low for an individual stock (particularly a small one), for which asymmetric information plays a major role, but is much too high for US government bonds. Breen et al. (2002) report average impact coefficients in the 20 to 30 range (in elasticity units) for

[11]These are typical of the parameter values used in Menzly et al. (2004) to describe the evolution of industry shares of the aggregate market. Santos and Veronesi (2005) use similar values for innovations to the share of dividends in all wealth (i.e., financial and human capital).

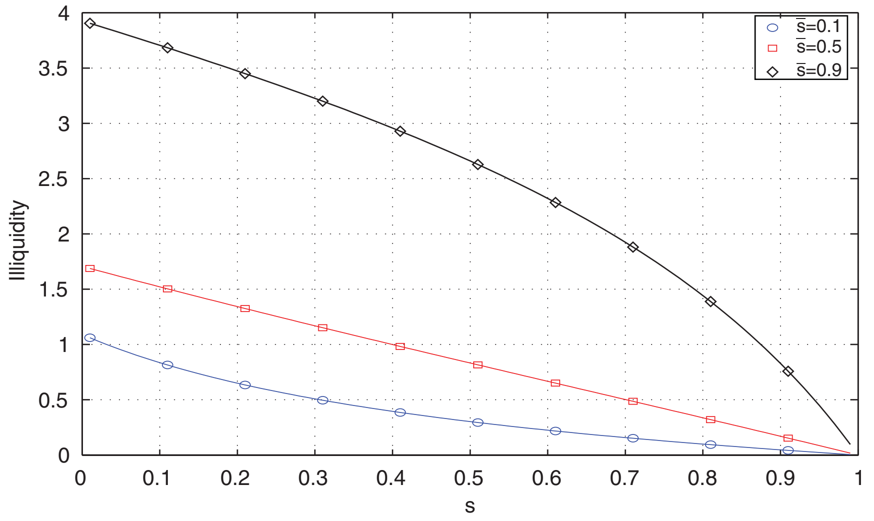

Fig. 2. The figure shows the illiquidity $\mathscr{I}$ as a function of s the consumption share of asset one, computed under the model of Santos and Veronesi (2005), c.f., Eqs. (6 and 7). The three curves all use the utility parameters $\phi = 0.01$ and $\gamma = 2$, and the endowment process parameters $\sigma_C = 0.03$, $\bar{\mu} = 0.03$, $\rho = 0.2$, $a = 0.2$ and $b = 0.6$. The long-run consumption share $\bar{s}$ varies as shown in the legend.

NYSE/AMEX stocks. On the other hand, the results of Glosten and Harris (1988) suggest that less than one quarter of that is likely to be permanent. Larger value transactions produce lower estimates. Loderer et al. (1991) report a median elasticity for seasoned equity offerings (SEOs) after controlling for information effects of approximately four. Studies of tender offers (Bagwell, 1992) and index alterations (Shleifer, 1986) find numbers below one. For Treasury bonds, Fleming (2001) reports a price impact coefficient of about 0.0005 for a \$10 million trade in two-year notes. The typical supply of such a note is \$10 billion and a typical price is 100. So this equates to an elasticity of just 0.005. Even during the LTCM crisis, this number never exceeded 0.02. While direct measurements of elasticities for broad market claims have, to my knowledge, not yet been reported, the relevant range is likely to be between these two sets of numbers. Hence, any model that implies $\mathscr{I}$ values from 0.01 to 1 may account for a large component of observed aggregate illiquidity.

The next-most important feature of the plots is that illiquidity goes up as s goes down, consistent with the analytical results above. Rising illiquidity in falling markets would seem to be an important feature for any dynamic model as it is consistent with both the empirical literature and everyday experience. The fact that it can be at least partially explained by a simple, frictionless representative agent model is surprising. It suggests that the formulation of illiquidity studied here is more than just a mathematical construct.

Fig. 3 again plots illiquidity against s, this time fixing $\bar{s} = 0.8$ and varying a. The intuitive logic above suggests that it is the mean-reversion in dividend shares that drives illiquidity in this model, with a determining the degree of mean-reversion. Here, that conjecture is confirmed graphically, with $\mathscr{I}$ driven towards zero as a declines. The attenuation is slow, however. If the characteristic time scale is of the order of ten years (the middle line), illiquidity is only about 20% lower at the steady-state value of s than if the time scale were one year (top line).

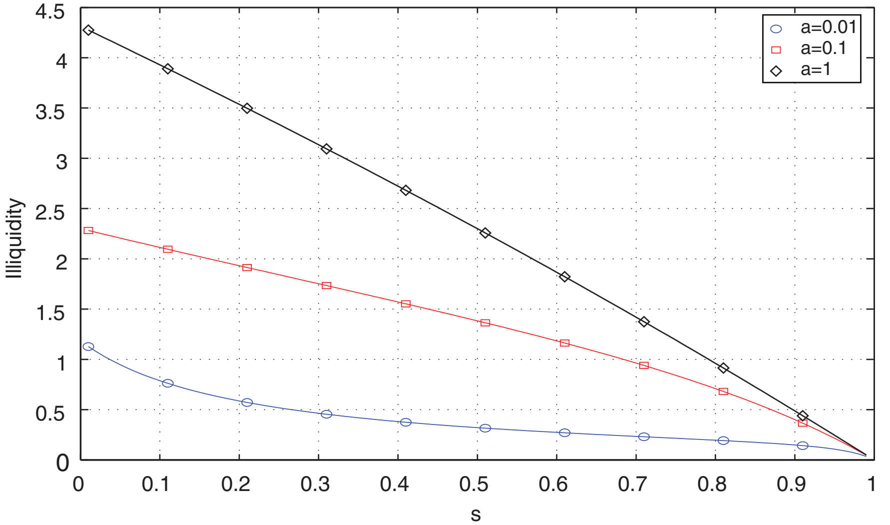

Fig. 3. The figure shows the illiquidity $\mathcal{I}$ as a function of s the consumption share of asset one, computed under the model of Santos and Veronesi (2005), c.f., Eqs. (6 and 7). The three curves all use the utility parameters $\phi = 0.01$ and $\gamma = 2$, and the endowment process parameters $\sigma_C = 0.03, \bar{\mu} = 0.03, \rho = 0.2, \bar{s} = 0.8$ and $b = 0.6$. The mean-reversion speed a varies as shown in the legend.

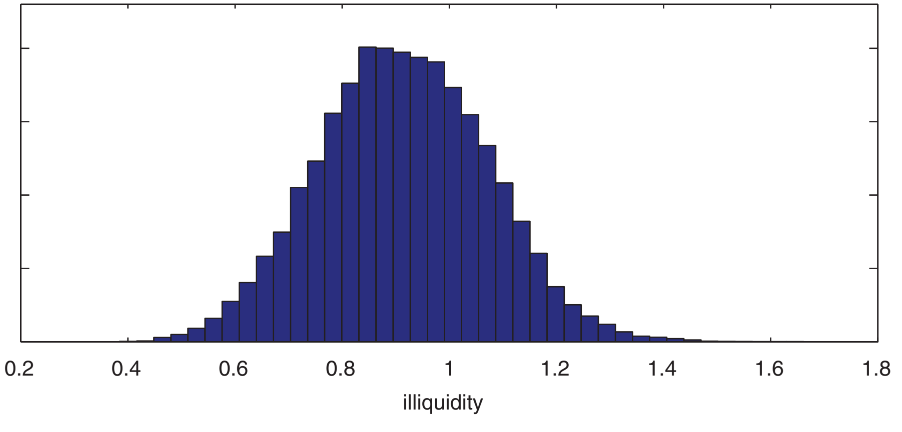

Fig. 4. The histogram shows the realized distribution of the illiquidity $\mathcal{I}$ for a 500-year simulation of the Santos and Veronesi (2005) model using the preference parameters $\phi = 0.01$ and $\gamma = 2$ and endowment process parameters $\sigma_C = 0.03, \bar{\mu} = 0.03, \rho = 0.2, \bar{s} = 0.8, a = 0.2$, and $b = 0.6$.

Now consider the covariation in liquidity with other barometers of the state of the economy. For the base-case parameter values, the stationary distribution of illiquidity is shown in Fig. 4. The distribution has a mean of 0.87 with a standard deviation of 0.17. This is the same order of variability that Fujimoto (2004) finds using the proportional spread measure of Eckbo and Norli (2002), and is only slightly less than that exhibited by

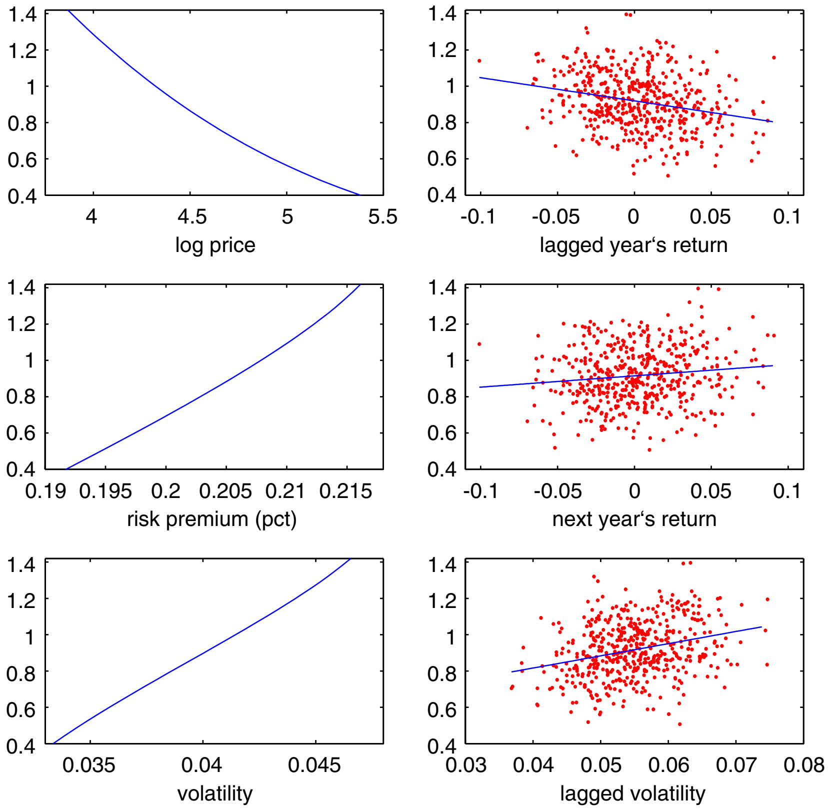

Fig. 5. The plots show the theoretical and simulated dynamic properties of the illiquidity $\mathscr{I}$ (on the vertical axes) versus other moments of the asset price process under the Santos and Veronesi (2005) model. The left hand figures show the exact functional relations between $\mathscr{I}$ and the log price level (top), the instantaneous expected excess return (middle), and the instantaneous volatility (bottom). The right panels plot the realized value of $\mathscr{I}$ versus realizations of lagged returns (top), future returns (middle), and lagged volatility (bottom) for the same model simulated over 500 years and sampled yearly. All panels employ the preference parameters $\phi = 0.01$ and $\gamma = 2$ and endowment process parameters $\sigma_C = 0.03, \bar{\mu} = 0.03, \rho = 0.2, \bar{s} = 0.8, a = 0.2$, and $b = 0.6$.

the measures of Jones (2002) and Amihud (2002). The half-life of liquidity shocks in this economy—indeed, the half-life of all shocks in this economy—is controlled by the parameter a, here set at 0.2. This makes for something of a compromise: empirical measures yield liquidity half-lives of roughly one year, whereas macroeconomic quantities (consumption shares, interest rates, etc.) have much slower decay.

Turning to the covariations, the left-hand panels of Fig. 5 plot the relation between illiquidity and, respectively, the price level, the expected excess return, and the instantaneous volatility of asset one. The top panel uses the (log) per-share price of the asset (holding the dividend of the other asset fixed) instead of s to show the degree to which $\mathscr{I}$ responds to the market level. The next two panels depict monotonically increasing relations between the first and second moments of instantaneous returns and illiquidity over the relevant range. Both risk premia and volatility are, of course, too small in this economy, since it is a consumption-based model with low risk aversion. However, there is still dynamic variation in these quantities, which correspond to significant changes in liquidity.

Verifying this, the right panels of Fig. 5 show the empirical relations, from a 500-year simulation of the model that correspond to the theoretical ones on the left. The top panel plots $\mathscr{I}$ against the previous year's excess returns, showing that high levels of aggregate illiquidity are expected after bad realizations, consistent with Chordia et al. (2000) and Huberman and Halka (2001). Using the impulse response measures computed by Fujimoto (2004), a one standard deviation positive shock to the stock market corresponds to a contemporaneous increase in liquidity of about 0.2 standard deviations. In the simulations, the response (in the same units) is 0.21. Likewise, a one standard deviation increase in volatility corresponds to a decrease of roughly 0.1 standard deviations in liquidity in the data. In the simulations, the response (measured by the fitted line in the bottom right-hand panel) is somewhat stronger at 0.32.

Most interestingly, the middle panel confirms that, conditional on high illiquidity, future excess returns are expected to be higher. This forecasting relation is found by Jones (2002) and Amihud (2002), whose numbers suggest that a one standard deviation shock to illiquidity raises annual expected returns by 2–4%. In the model, the risk premium variation is nowhere near this level because the premia are so low. However, the association is still statistically strong. An observer of this economy could well conclude that liquidity risk is priced, even though it is not.

Finally, this model can be used to study the relation between market liquidity and the liquidity of its individual components.[12] Suppose that "the market" actually consists of N sectors or firms, each of whose consumption share obeys a multivariate analogue of Eq. (7).[13] Then the sum of all N assets also has the same dynamic form, and the liquidity of any one stock will depend on both its share, $s^{(i)}$, and the market's share, $s^{(M)} = 1 - s^{(0)}$. Specializing to the case of log utility and equally risky firms, the relation between consumption shares and liquidities can be analyzed explicitly.

Proposition 3.2. *Assume the conditions of Proposition* 3 *in* Santos and Veronesi (2005). *Then*

(i) $\frac{\mathrm{d}\mathscr{I}^{(M)}}{\mathrm{d}s^{(M)}}(s^{(M)})<0$,
(ii) $\frac{\partial\mathscr{I}^{(i)}}{\partial s^{(M)}}(s^{(i)}, s^{(M)})<0, \quad \forall\, s^{(i)}$, *and*
(iii) $\frac{\mathrm{d}\mathscr{I}^{(i)}}{\mathrm{d}s^{(M)}}(\lambda\, s^{(M)}, s^{(M)})<0, \quad \forall\, \lambda$.

The first equation confirms the results above: liquidity falls in down markets. (Note that it is easy to show that $P^{(M)}$ is a monotonic function of $s^{(M)}$. So the definition of "down markets" is unambiguous.) The next two show that individual stock's liquidities also fall with the market, whether their individual dividends are fixed (second line) or whether $s^{(i)}$ falls in proportion to $s^{(M)}$ (third line). Putting the conclusions together, it is also clear that individual liquidities must be monotonically related to the market liquidity, and thus will inherit the dynamic properties found above.

[12] I thank the referee for suggesting the following analysis.

[13] The exact specification is $\mathrm{d}s_i = a(\bar{s}_i - s_i)\,\mathrm{d}t + s_i\sigma_i(s)\,\mathrm{d}B_t$, where B is a vector of standard Brownian motions, $\sigma_i(s) = (v_i - \sum_{j=0}^{N} s_j v_j)$, and the v_i are (row) vectors of constants that determine the covariance structure.

All together this model seems to deliver a remarkably accurate set of predictions about both the levels and the dynamic properties of liquidity. Even though it is not at all designed to say anything about liquidity, it does embed a nontrivial explanation for supply effects: if a supply shock does not alter the long-run equilibrium relations in the economy, the path back to that equilibrium will involve an erosion of the immediate increase in gross dividends. The results here suggest that this mechanism may, in fact, account for an important component of aggregate liquidity.

3.2. *The model of Cochrane, Longstaff, and Santa Clara (2004)*

A reasonable question about the results above is whether the assumption of a cointegrated dividend process is necessary to generate liquidity effects. The example in this section will show that this is not the case.

Cochrane et al. (2004) (hereafter, CLSC) study an economy similar to that of Santos and Veronesi (2005) except that the individual dividend processes are not linked by any long-run relation. Again, there are two assets on an infinite horizon with power utility.[14] But now each payout stream is a geometric Brownian motion. These are allowed to be correlated. Even so, the share of each dividend in consumption will diverge almost surely to zero or one as time increases. So the economy is nonstationary, and all the dynamic features are transitory. Still, the model describes a situation that may be more relevant for the case of individual companies, in which long-run survival is not guaranteed.

The notation is basically the same as above. The per-share dividend of each asset now obeys

$$\frac{\mathrm{d}D^{(i)}}{D^{(i)}} = \mu^{(i)}\,\mathrm{d}t + \sigma^{(i)}\,\mathrm{d}W_t^{(i)},$$

where the coefficients are constant. Here ρ will denote the correlation (also constant) between $\mathrm{d}W^{(0)}$ and $\mathrm{d}W^{(1)}$. Again, total consumption is $C_t = D_t^{(0)}X^{(0)} + D_t^{(1)}X^{(1)}$. And again the key state variable is $s_t \equiv D_t^{(1)}X^{(1)}/C_t$.

In this model s matters because it determines instantaneous risk. The higher s, the more asset one's dividend shocks affect consumption. Hence, the more risk premium is attached to its future dividends. When $D^{(1)}$ is very small, the correlation between $\mathrm{d}D^{(1)}$ and $\mathrm{d}C$ is negligible, and thus cash flows to this asset are discounted roughly at the riskless rate. But as $D^{(1)}$ grows, its discount rate also rises, and this dampens the price response to dividend shocks.

This simple dynamic also explains why illiquidity arises in this setting. Adding shares of asset one *also* raises the risk of that asset. Hence, discount rates go up and prices (of old and new shares) fall. In essence, there is a negative externality from the additional supply.[15]

Defining the price–dividend ratio via $P^{(1)}\ X^{(1)} \equiv D^{(1)}\ g(s)$ and the wealth-consumption ratio via $P^{(0)}X^{(0)} + P^{(1)}X^{(1)} \equiv C\ F(s)$, the functions $g(s)$ and $F(s)$ can be found by solving relatively simple one-dimensional boundary value problems. (At the extreme points, $s = 0$ and $s = 1$, the economy collapses to the standard Lucas, 1978 case.) Given these, the

[14] CLSC actually impose log utility in order to obtain semi-closed-form results. Without this assumption it is still easy solve the model's pricing equation numerically.

[15] This discount rate effect is also present in the Santos and Veronesi (2005) model. But there it matters much less because the long-run level of risk is fixed by $\bar{s}$.

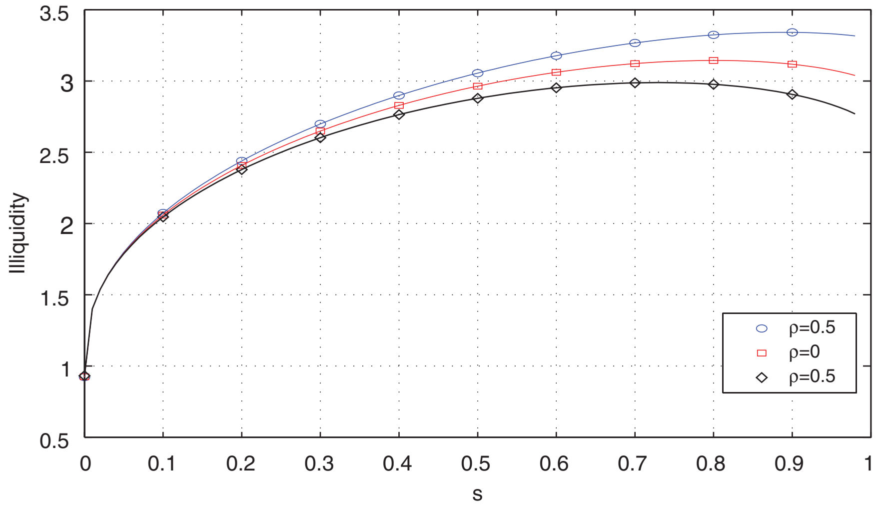

Fig. 6. The figure shows the elasticity $\mathscr{I}$ as a function of s, the consumption share of asset one under the model of Cochrane and Longstaff (2004) for the case of log utility with subjective discount factor $\phi = 0.01$. The parameter settings for the two endowment processes are $\mu_1 = 0.10,\ \sigma_1 = 0.20$ and $\mu_0 = 0.03,\ \sigma_0 = 0.03$, with the correlation ρ varying as shown in the legend.

computation of $\mathscr{I}$ is straightforward, since, as shown in the proof of Proposition 3.1, in any model for which prices are only a function of s, we have

$$\mathscr{I} = -s(1-s)\,\frac{g'(s)}{g(s)}\left(\frac{F(s)}{F(s) - s\,g(s)}\right). \tag{10}$$

Fig. 6 shows $\mathscr{I}$ as a function of s in the log utility case. The parameterization uses $\mu_1 = 0.10, \sigma_1 = 0.20, \mu_0 = 0.03$, and $\sigma_0 = 0.03$. For small values of s, this case can be interpreted as portraying the dynamics of an individual stock (or a sector) versus a claim with the characteristics of aggregate consumption (i.e., "the market" in a typical endowment economy calibration). For larger s, the case might be better viewed as describing the returns on levered equity (asset one) versus nominal government bonds (asset zero).[16]

There are actually two mechanisms through which supply shocks affect the risk premium on asset one. As already described, the correlation of the more risky asset's dividend with consumption rises with s, and this effect is stronger for larger s. Moreover, the volatility of consumption itself changes. If $\rho > 0$ this works in the same direction, whereas for $\rho < 0$ it can work the opposite way, lowering the risk premium as s increases. The risk premium dynamics feed directly through to illiquidity, as the figure illustrates.

[16]These parameter choices (and the resulting dynamics) are similar to those of the "bonds-stock" case examined by CLSC. I also follow the original paper in focusing on log utility, but continue to employ the more standard subjective discount rate $\phi = 0.01$ rather than their 0.10.

Looking at the plots, it is again worth pointing out the most basic features of liquidity in this example: both its level and variability are nontrivial. Thus, the CLSC model also provides a potential explanation for the existence of systematic liquidity effects.

Unlike the Santos and Veronesi (2005) model, this one does not, in general, imply that liquidity goes up as an asset's share of the economy increases. On the other hand, the *price* of the asset will increase with positive shocks to D_1 or to D_0. The former will raise s, in turn raising $\mathscr{I}$ as in the plot. But the latter will lower s, leading to the empirically observed positive relation between prices and liquidity. If we view asset one as a small stock (with s close to zero) and asset zero as the market, then the theory yields the interesting testable implication that liquidity should covary in opposite directions with idiosyncratic and systematic return shocks. However, since own-dividend shocks are more volatile and have a bigger impact on prices, the unconditional return-liquidity effect will apparently go the wrong way.

To further explore the dynamics, Fig. 7 shows the relations between $\mathscr{I}$ and the other characteristic quantities for the same parameterization.[17] Here I focus on the behavior that would characterize a small stock, restricting attention to the range $s < 0.15$. The panels on the right show the functional relations with (reading down) the price level, the instantaneous expected return, and the instantaneous volatility. I also plot the corresponding relations between $\mathscr{I}$ for the stock and the *market's* level, expected return, and volatility (shown as dashed lines on the same axes), where the market is defined as the total wealth portfolio comprising both assets. The bottom two panels, like the model in the last section, present the desired positive associations. Illiquidity is higher both when expected returns and volatility are higher for the stock itself and when they are higher for the market. The dashed line in the top panel shows that, holding this asset's dividends fixed, higher illiquidity also corresponds to times when the market overall is down. The solid line in this panel shows the problematic positive association between illiquidity and the asset's own level (holding the other asset's payout fixed).

To check how these theoretical relations map into sample properties, I simulate 10 paths of 50 years of the economy with s starting at 0.001 to get a sense of what the stochastic implications for small s would be. (With this starting value, s always stays between zero and 0.15, which is the range shown in the left-hand plots.) The results are not meaningful in a statistical sense because the economy is nonstationary. Nonetheless, they give some idea of the model's likely ability to describe real data.

The panels on the right side of Fig. 7 show the output of this exercise using returns and volatilities tabulated at one-year intervals. In the top panel, the theoretical association between the overall market level and the small stock's illiquidity turns out to be imperceptible because, here the market impounds both the effects of D_1 and D_0 shocks, which affect illiquidity in opposite directions. Similarly, in the middle panel, the ability of $\mathscr{I}$ to forecast returns is undetectable in practice, which is due to the vary small magnitudes of the risk premia under log utility. Finally, the bottom panel offers a similar conclusion about the volatility–liquidity relation in this model. The theoretically correct relation is hard to pick up when the variation in volatility is so small.

Of course, these failings are mostly a consequence of the highly stylized nature of the model which is designed to illustrate a particular equilibrium mechanism, and not to

[17]Following CLSC, I put $\rho = 0$ here. This has a negligible effect over the s range analyzed, as the previous plot suggests.

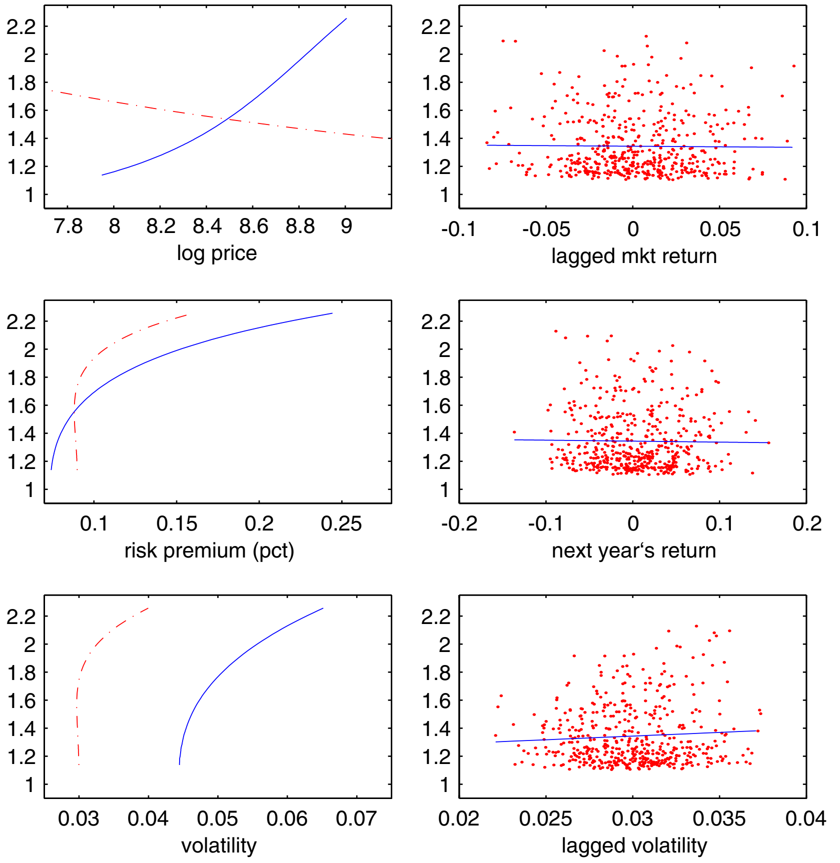

Fig. 7. The plots show the theoretical and simulated dynamic properties of the illiquidity $\mathscr{I}$ (on the vertical axes) versus other moments of the asset price process under the Cochrane and Longstaff (2004) model. The left hand figures show the exact functional relations between $\mathscr{I}$ and the log price level (top), the instantaneous expected excess return (middle), and the instantaneous volatility (bottom). The right panels plot the realized value of $\mathscr{I}$ versus realizations of lagged aggregate returns (top), future own returns (middle), and lagged own volatility (bottom) for the same model simulated over 500 years and sampled yearly. All panels employ the preference parameters $\phi = 0.01$ and $\gamma = 1$ and endowment process parameters $\mu_1 = 0.10$, $\sigma_1 = 0.20$ and $\mu_0 = 0.03$, $\sigma_0 = 0.03$, and $\rho = 0.0$.

address the well known empirical issues with consumption based models. Its shortcomings should not detract from the basic observation that the model delivers its own theory of liquidity that is distinct from that of Santos and Veronesi (2005). It illustrates another important channel—a risk exposure effect—for endogenous liquidity determination, and shows that this effect implies both economically meaningful levels of illiquidity and new dynamic predictions, which are testable.

3.3. The Lucas (1978) model revisited

As a last example, I compute illiquidity in the classic Lucas (1978) model to demonstrate a third mechanism that gives rise to illiquidity in equilibrium, namely, a risk-free rate effect. An increase in the relative supply of the risky asset makes aggregate consumption more volatile, which can lead to an increase in precautionary savings. This lowers the riskless rate, making all assets more valuable. This works against the risk premium effect of the previous example and lowers illiquidity as s rises.

A second motivation for studying this model is to underscore the fact that the *real* liquidity effects quantified by $\mathscr{I}$ are not somehow undone when prices are measured in units of the exchange asset. In this model, asset zero is an instantaneously riskless "money market account" whose price is identically one. Hence, real and nominal effects are the same. In the two previous models, the "safe" asset still has a price that varies with the state of the economy. (This is even true in the CLSC model with $\sigma_0 = 0$.) So it is reasonable to ask whether that variability is responsible for the dynamics of $\mathscr{I}$. Here we will dispel that possibility.

Finally, the Lucas (1978) model allows me to illustrate another conceptual point about the notion of liquidity introduced here: it can be computed for assets even if they are in zero supply. In the standard formulation of this model, the supply of the riskless asset is set to zero as a condition of the equilibrium. However, there is no reason why the elasticity $\mathscr{I}$ cannot be computed by varying this supply away from zero. As we will see, this leads to a nontrivial limit. This same technique could just as well be used to analyze liquidity in derivatives or insurance markets.

To carry out the calculation, we need to solve the generalized Lucas (1978) model for the case in which $X^{(0)}$ is not zero. In fact, we will then be able to study $\mathscr{I}$ for *any* value of the riskless asset supply, not just the limit at zero. This will yield a third dynamic economy since, fixing $X^{(0)} > 0$, the consumption share of the risky asset s will again vary with stochastic changes in the two dividend streams.

Note that now the riskless rate r_t must be determined in equilibrium to enforce the condition that the price of the riskless asset is constant.[18] With this stipulation, the riskless rate is determined by the usual Euler equation, which dictates that this rate is the (negative) drift of the stochastic discount factor. Marginal utility, however, now depends on r_t since this is a component of consumption. So this leads to a differential equation for r as a function of D_t, $X^{(0)}$, and $X^{(1)}$.

Letting $x = D_t X^{(1)}/X^{(0)}$, this equation turns out to be

$$\tfrac{1}{2}\gamma(1+\gamma)\sigma^2 x^2[\dot{r}+1]^2 - \gamma(r+x)[\mu x(\dot{r}+1) + \tfrac{1}{2}\sigma^2 x^2 \ddot{r}] + (r-\phi)[r+x]^2 = 0, \tag{11}$$

where $\dot{r} = \mathrm{d}r/\mathrm{d}x$ and $\ddot{r} = \mathrm{d}^2 r/\mathrm{d}x^2$.

Solving this determines the consumption process, and hence the risk premium on the risky asset. Let P denote a claim to the normalized dividend flow x. Then its Euler equation yields the differential equation:

$$\frac{1}{2}\sigma^2 x^2 \ddot{P} + \left[\mu x - \gamma\sigma^2 x^2\left(\frac{\dot{r}+1}{r+x}\right)\right]\dot{P} - rP + x = 0. \tag{12}$$

Once P is obtained, it is straightforward to compute $\mathscr{I}$: simply differentiate $P^{(1)} = (X^{(0)}/X^{(1)})\,P(x)$ with respect to $X^{(1)}$ along the value isoquant given by $\mathrm{d}X^{(0)}/\mathrm{d}X^{(1)} = -P^{(1)}/P^{(0)} = -P^{(1)}$.

[18]Since its dividends are endogenous, the riskless asset cannot now be viewed as a primitive technology. The mechanism underpinning this could be a commitment by a government outside the model to pay (or receive, if $X^{(0)} < 0$) the variable rate r_t, set according to the solution below.

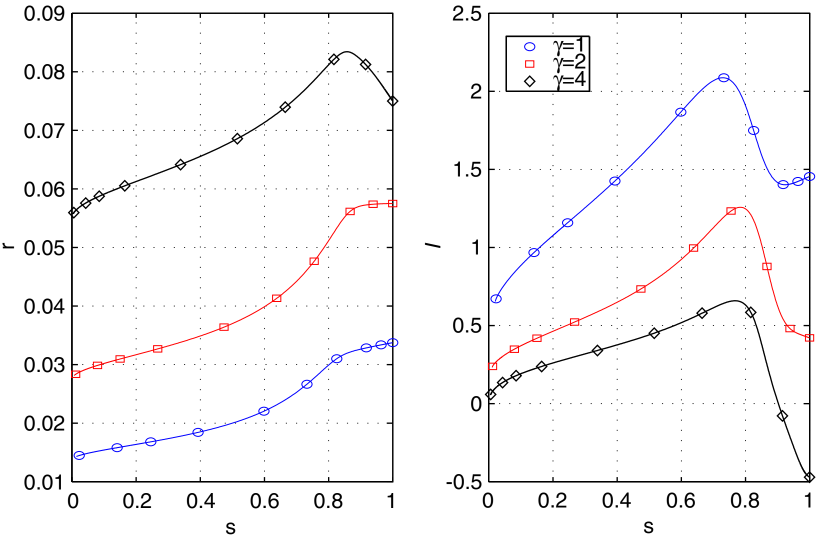

Fig. 8. The figure shows the riskless rate, r, and the elasticity, $\mathcal{I}$, as a function of the risky asset's endowment share, s, under the generalized Lucas (1978) model for different values of the preference parameters γ and ϕ. The lines labelled with circles use $\gamma = 1$, $\phi = 0.01375$; those labelled with squares use $\gamma = 2$, $\phi = 0.0275$; and those labelled with diamonds use $\gamma = 4$, $\phi = 0.055$. The three curves all have the endowment growth rate and volatility set at $\mu = 0.03$ and $\sigma = 0.10$, respectively.

There are complications in solving both of the differential equations, and, in fact, neither one obeys standard conditions for existence and uniqueness of solutions. Under some restrictions on the parameters, one can compute solutions on the interior of the unit interval for $w = 1/(1 + x)$ and show continuity on the boundaries. Since the boundary cases correspond to single asset economies, the limit values of r and P (or $\dot{P}$) are known. Having solved for P, r, and $\mathcal{I}$ as a function of x, we can then view these as functions of the consumption share of the risky asset (since $s = x/(x + r)$), for comparison with the results in the previous subsections.

Fig. 8 shows the riskless rate and illiquidity for three parameter configurations using risk aversion coefficients of one, two, and four.[19] As with the previous examples, the modified Lucas (1978) model also delivers economically significant levels of liquidity, with its own characteristic patterns. All cases show the general pattern found above in the CLSC model: $\mathcal{I}$ rising with s due to the risk premium effect. For large s, however, the precautionary savings effect causes r to drop with s, leading to increases in liquidity.[20] In the most risk-averse case, $\mathcal{I}$ can even become negative. This can be interpreted logically as a situation in which the representative agent would be a net demander of liquidity.

[19]Here the subjective discount rate increases (for numerical reasons) from 0.01375 to 0.0275 to 0.055 across the cases, which simply shifts the curves vertically. The solutions also use $\mu = 0.03$ and $\sigma = 0.10$. The higher endowment volatility is also chosen for computational convenience, but is not out of the range of plausible values when calibrated to dividends.

[20]The CLSC model shows no such effect for the log utility case, in which r is constant, but it has the same pattern for cases with $\gamma > 1$.

Another interesting observation from the graphs is that liquidity is not infinite in the Lucas (1978) case, when the riskless asset is in net zero supply. In fact, the limiting value of $\mathscr{I}$ at $s=1$ (the right-most endpoints of the lines in the left panel) can be found analytically.

Proposition 3.3. *Assume the differential Eqs.* (11) *and* (12) *have solutions with finite derivatives as $x \to \infty$. Then $\mathscr{I}(x)$ approaches the limit*

$$q_0(\mu\gamma - \sigma^2\gamma^2), \tag{13}$$

where

$$q_0 = 1/(\phi - \mu(1-\gamma) + \tfrac{1}{2}\gamma(1-\gamma)\sigma^2)$$

is the Lucas price–dividend ratio.

This value is positive whenever $\mu/\gamma > \sigma^2$, which is satisfied for reasonable dividend processes unless risk aversion is very large. For typical values of σ, the function decreases with γ, which, again, is a consequence of the risk-free rate effect.

When $X^{(0)}$ is not zero, liquidity is dynamic and one can examine its variation with other quantities, as with the preceding models. As in the CLSC model, the evolution is transitory, and covariances do not exist. However, the functional instantaneous relations can still be computed. Fig. 9 plots $\mathscr{I}$ against the level of the market, the risky asset's volatility, the price–dividend ratio, and the risk premium for the log utility case.

All four plots describe nonmonotonic relations with two inflection points. The share of the risky asset rises going from left to right in each plot, with the first inflection occurring at about $s=0.75$ and the second at $s=0.91$. When the riskless asset is large ($s<0.75$), the relations look encouraging: illiquidity rises with volatility and risk, as shown in the right-hand panels. The same holds when the risky asset is large, although the range in $\mathscr{I}$ is very small when $s>0.91$. As with the CLSC model, however, the theoretically correct relations are essentially impossible to distinguish in simulated data due to the small range of volatilities and risk premia.

Again, the purpose at present is not to offer a particular model that accurately accounts for all of the fluctuations in aggregate liquidity. Instead, it is to illustrate a range of possible theories and to explain their economic intuition. The riskless rate effect, the risk premium effect, and the dividend cointegration effect each imply nontrivial price responses to perturbations in quantities. Hence, an investor needing to trade a non-infinitesimal quantity, for whatever reason, would encounter a degree of illiquidity that is determined by the underlying state of the economy.

4. Conclusion

This paper contributes to the growing branch of research on time-varying market liquidity. However, instead of viewing liquidity as exogenous and deduce the implications for equilibrium returns, I reverse the logic and deduce endogenous implications about liquidity from existing equilibrium models. This inquiry is important for three reasons.

First, it has not been previously recognized that such latent liquidity effects in fact exist in frictionless models. Building on the insights of classical inventory models in microstructure, I quantify the willingness of a representative agent—who here plays the role of the market maker—to accommodate marginal trades that would alter his portfolio. In general, such trades lead to nontrivial supply effects in endowment economies because

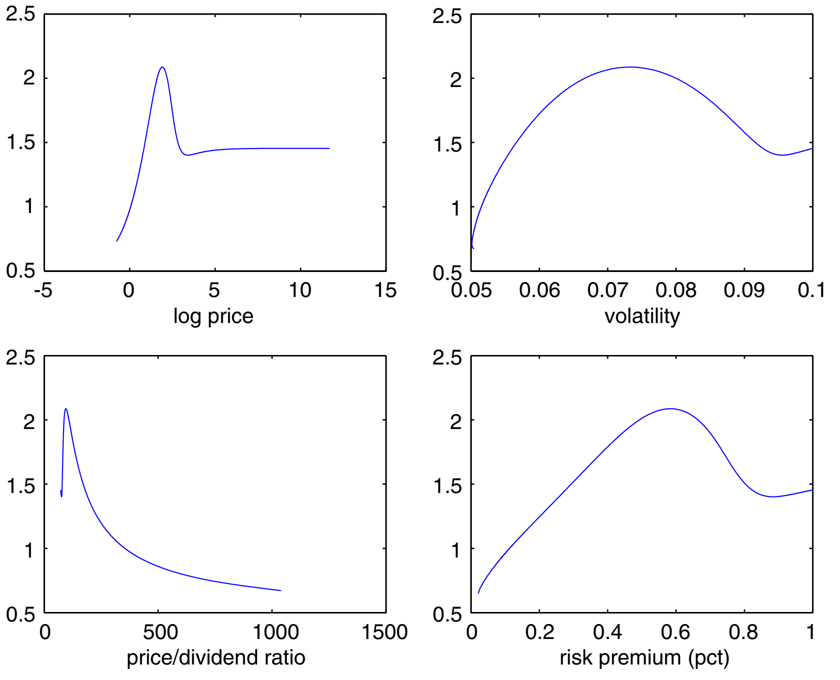

Fig. 9. The figure shows the illiquidity, $\mathscr{I}$, as a function of four dynamic quantities characterizing the risky asset's price process in the generalized Lucas (1978) model. The horizontal axes, going clockwise from the upper left, are the price level of the risky asset, its volatility, its expected excess return, and the price–dividend ratio. The preference parameters are $\gamma = 1$ and $\phi = 0.01375$, and the endowment process parameters are $\mu = 0.03$ and $\sigma = 0.10$.

they (marginally) change the dynamics of aggregate consumption. This notion of liquidity is a property of the equilibrium itself and has nothing to do with the institutional details of the exchange process. While I view the costs as hypothetical in the same sense that prices are hypothetical (where there is no actual trade), I also show that my measure of illiquidity corresponds to the price impact or bid–ask spread that would be faced by agents who do trade for unmodelled reasons.

Second, the asset pricing consequences of liquidity risk cannot be fully assessed until the exogenous and endogenous determinants of liquidity are disentangled. It only makes sense to regard liquidity as a risk factor if it is not determined by other modelled state variables. Undoubtedly the frictionless models studied here omit other important determinants of liquidity. However, if the endogenous component is indeed significant, that may restrict the role of exogenous factors. As a practical matter, if liquidity risk is spanned by other variables, it can be hedged with existing assets.

Third, market liquidity is an important quantity in its own right that should be explained endogenously. It has a number of intriguing dynamic properties: it covaries with volatility and prices; and it has forecasting power for future returns. These facts constitute largely unexplored territory for theorists. Stepping back from asset pricing questions, I suggest that we can and should regard equilibrium models as offering falsifiable explanations for these aggregate effects.

To demonstrate this, I compute the illiquidity measure $\mathscr{I}$ in three example economies in which liquidity varies for different reasons. None of the models was originally intended to explain liquidity. Yet, remarkably, the parameterizations studied do, indeed, get a lot of the qualitative features right. With plausible fundamental parameters, $\mathscr{I}$ is economically large, and can exhibit a positive correlation with contemporaneous returns and a negative relation with volatility and expected returns. In the case of the Santos and Veronesi (2005) model, in fact, the magnitudes of the dynamic responses in simulated data accord well with the same quantities measured empirically. In general, the theoretical relations can be nonmonotonic and quite complex, which may help to explain why liquidity sometimes seems like a distinct state variable.

While the results suggest that, to some degree, changes in liquidity may be driven by the underlying risk dynamics of the economy, this is only an initial step. The models I examine in Section 3 fail along a number of dimensions empirically. (Both volatilities and expected returns are too low and too stable, for example.) So it is unrealistic to expect them to provide a full account of liquidity dynamics. Nonetheless, the approach introduced here points the way towards the development of a fuller description.

Some natural generalizations of the examples immediately suggest themselves. Models with an explicit role for fiat money (e.g., via cash-in-advance constraints) would enable the computation of elasticities with respect to it, which may be the most empirically relevant quantities. Such models could also delineate a connection between market liquidity and monetary liquidity or credit conditions, which is completely independent of any frictions in financial intermediation.

Similarly, models involving more realistic pricing kernels would undoubtedly induce more realistic liquidity. Models with time-varying fundamental risk (stochastic volatility) or risk aversion (habits), in which expected returns vary rapidly, would imply more variable liquidity. Any model with higher market risk premia could produce the appearance of a larger price of liquidity risk.

In addition, models with two or more risky assets (besides the numeraire) would permit examination of the endogenous evolution of relative liquidity, such as contagion and flight-to-quality effects. Again, the approach would illuminate an aspect of these effects that is implicit in existing representative agent models, and that is thus not driven by other information effects, irrationality, or market imperfections.

Finally, the approach here would be further enhanced by the incorporation of whatever frictions might give rise to demand for liquidity (or shocks to asset supplies). The elasticity $\mathscr{I}$ can be computed just as well in models with solvency constraints or taste shocks or other mechanisms yielding explicit trade. This could shed light on the interaction between liquidity and volume or flows.

Appendix: Proofs

This appendix collects proofs of the results in the text.

Proof of Proposition 2.1

To start, I impose the restriction that the mapping $\Theta(x) = \Theta(x; X; \mathscr{D})$ defined in Eq. (2.1) exists and has a continuous first derivative.

Now since the liquidity demanders are only permitted to alter their holdings at the trade dates, the holdings of the representative agent $\widehat{\mathscr{R}}$ are also fixed between these dates.

Consider the market at the final trade date, τ_K. At that point, the value function of $\widehat{\mathscr{R}}$ if he does not buy shares is

$$\mathrm{E}_{\tau_K}\left[\int_{\tau_K}^{T} \mathrm{e}^{-\phi(s-t)} u\left(\sum_{i=0}^{N} \widehat{X}^{(i)}_{\tau_{K-1}} D^{(i)}_s\right) \mathrm{d}s\right],$$

which is $J(\widehat{X}_{\tau_{K-1}}; \tau_K; \mathscr{D}_{\tau_K})$ as defined in the proposition. If $\widehat{\mathscr{R}}$ does purchase the shares, paying price Δ in units of asset zero, his value function is

$$J(\widehat{X}^{(0)}_{\tau_{K-1}} - \Delta, \widehat{X}^{(1)}_{\tau_{K-1}} + \varepsilon, \widehat{X}^{(2)}_{\tau_{K-1}}, \ldots, \widehat{X}^{(N)}_{\tau_{K-1}}; \tau_K; \mathscr{D}_{\tau_K}).$$

Call this $\widehat{J}(\Delta)$. (Its dependence on the portfolio holdings may include any changes induced in endogenous state variables.)

Under perfect competition, $\widehat{\mathscr{R}}$'s optimal bid, B, will satisfy

$$B = \left\{\inf_{\Delta} : \widehat{J}(\Delta) \geqslant J(\widehat{X}_{\tau_{K-1}}; \tau_K; \mathscr{D}_{\tau_K})\right\}.$$

The assumption above on the mapping Θ is sufficient to guarantee that the infimum here always achieves equality. Hence, the agent bids so as to remain on the value isoquant. Specifically,

$$B = \Theta(\widehat{X}^{(1)}_{\tau_{K-1}} + \varepsilon; \widehat{X}_{\tau_{K-1}}, \mathscr{D}_{\tau_K}) - \widehat{X}^{(0)}_{\tau_{K-1}}.$$

Similarly, the ask price is given by

$$A = \widehat{X}^{(0)}_{\tau_{K-1}} - \Theta(\widehat{X}^{(1)}_{\tau_{K-1}} - \varepsilon; \widehat{X}_{\tau_{K-1}}, \mathscr{D}_{\tau_K}).$$

Thus, whether shares are bought or sold in this auction, the value function will not change.

For any time after the next-to-last trade date, $\tau_{K-1} < t \leqslant \tau_K$, the value function is then

$$\begin{aligned}
&\mathrm{E}_t\left[\int_{t}^{\tau_K} \mathrm{e}^{-\phi(s-t)} u\left(\sum_{i=0}^{N} \widehat{X}^{(i)}_{\tau_{K-1}} D^{(i)}_s\right) \mathrm{d}s\right] + \mathrm{E}_t[J(\widehat{X}_{\tau_{K-1}}; \tau_{K-1}; \mathscr{D}_{\tau_K})] \\
&\quad = \mathrm{E}_t\left[\int_{t}^{T} \mathrm{e}^{-\phi(s-t)} u\left(\sum_{i=0}^{N} \widehat{X}^{(i)}_{\tau_{K-1}} D^{(i)}_s\right) \mathrm{d}s\right].
\end{aligned}$$

And this is the definition of the myopic function $J(\widehat{X}_t; t; \mathscr{D}_t)$.

By the same reasoning, J will be unchanged as a result of the trades at τ_{K-1} as well. Hence, for $\tau_{K-2} < t \leqslant \tau_{K-1}$, the value function is likewise $J(\widehat{X}_t; t; \mathscr{D}_t)$. The logic can now be repeated back to any arbitrary t, yielding the conclusion that the value function is the same function of $\widehat{X}^{(0)}$ and $\widehat{X}^{(1)}$ as it would be if these holdings were static for all times in the future. Since its partial derivatives with respect to these arguments must also be the same, the first statement in the proposition is established.

The agent's bid and ask at each trade date must therefore be the functions B and A deduced for the final trade date. The half-spread, $P^{(1,A)} - P^{(1)}$, in price units is thus

$$\begin{aligned}
&P^{(1)}(\widehat{X}^{(0)}_t - A, \widehat{X}^{(1)}_t - \varepsilon) - P^{(1)}(\widehat{X}^{(0)}_t, \widehat{X}^{(1)}_t) \\
&\quad = P^{(1)}(\Theta(\widehat{X}^{(1)}_t - \varepsilon), \widehat{X}^{(1)}_t - \varepsilon) - P^{(1)}(\widehat{X}^{(0)}_t, \widehat{X}^{(1)}_t).
\end{aligned}$$

Dividing by ε and letting ε approach zero yields the directional derivative $\mathrm{d}P^{(1)}(\Theta, \widehat{X}^{(1)})/\mathrm{d}\widehat{X}^{(1)}$. Normalizing this by $\widehat{X}^{(1)}/P^{(1)}$ and combining with the analogous limit for the bid side yields the conclusion of the proposition. $\square$

Proof of Proposition 2.2

At the first auction date, each agent has the same allocations x_0 of all shares. Hence, each has the same reservation values, B and A, for the additional ε shares. Since bidding up to the next-highest reservation value is the unique dominant strategy in an English auction, each agent will bid up to B and win shares. (And similarly for the offering auction.) Since there is a finite collection of auctions in total, their are no collusive equilibria, and playing the sequence of dominant strategies in individual auctions is the unique dominant strategy for the whole game. Hence, all agents act the same and their allocations stay equal. This yields the first conclusion (i) of the proposition.

Now the homogeneity of the agents and the distribution of shares implies that the economy as a whole can be characterized by a representative agent of the same preferences and beliefs. If the value function is also homogeneous in the position sizes, then the representative investor holding M times the position of each agent will have the same marginal utility ratios and hence prices and price elasticities.

In computing these, and establishing the myopic formulation of the value function, the argument follows the same logic as in the proof of the last proposition.

Since there are a finite number of trade dates, we start at the last one, τ_K. At that point, the value function of the representative agent if he does not buy shares is

$$\mathrm{E}_{\tau_K}\left[\int_{\tau_K}^{T} \mathrm{e}^{-\phi(s-t)} u\left(\sum_{i=0}^{N} X^{(i)}_{\tau_{K-1}} D^{(i)}_s\right)\mathrm{d}s\right] = J(X_{\tau_{K-1}}; \tau_K; \mathscr{D}_{\tau_K}).$$

If he does purchase the shares, paying price Δ in units of asset zero, his value function is

$$J(X^{(0)}_{\tau_{K-1}} - \Delta, X^{(1)}_{\tau_{K-1}} + \varepsilon, X^{(2)}_{\tau_{K-1}}, \ldots, X^{(N)}_{\tau_{K-1}}; \tau_K; \mathscr{D}_{\tau_K}) \equiv J(\Delta).$$

Since his optimal bid is his reservation value, B will satisfy

$$B = \left\{\inf_{\Delta} : J(\Delta) \geqslant J(X_{\tau_{K-1}}; \tau_K; \mathscr{D}_{\tau_K})\right\}.$$

We continue to assume the mapping Θ is sufficient to guarantee that the infimum achieves equality. Hence, the agent bids so as to remain on the value isoquant

$$B = \Theta(X^{(1)}_{\tau_{K-1}} + \varepsilon; X_{\tau_{K-1}}, \mathscr{D}_{\tau_K}) - X^{(0)}_{\tau_{K-1}}.$$

Likewise,

$$A = X^{(0)}_{\tau_{K-1}} - \Theta(X^{(1)}_{\tau_{K-1}} - \varepsilon; X_{\tau_{K-1}}, \mathscr{D}_{\tau_K}).$$

Because the value function will not change as a result of the trade, immediately before τ_K it is exactly the same function as it would be if the trade were not to take place. That is, it is equal to the myopic function $J(X_{\tau_K-}; \tau_K-; \mathscr{D}_{\tau_K-})$. Clearly the same conclusion applies for any $t \in (\tau_{K-1}, \tau_K]$. By backwards induction, the same logic holds for each trade date.

The agent's bid and ask at each trade date must therefore be the functions B and A deduced for the final trade date. The half-spread, $P^{(1,A)} - P^{(1)}$, in price units is thus

$$\begin{aligned} & P^{(1)}(X_t^{(0)} - A, X_t^{(1)} - \varepsilon) - P^{(1)}(X_t^{(0)}, X_t^{(1)}) \\ & \quad = P^{(1)}(\Theta(X_t^{(1)} - \varepsilon), X_t^{(1)} - \varepsilon) - P^{(1)}(X_t^{(0)}, X_t^{(1)}). \end{aligned}$$

Dividing by ε and letting ε approach zero yields the directional derivative $\mathrm{d}P^{(1)}(\Theta, X^{(1)})/\mathrm{d}X^{(1)}$. Normalizing this by $X^{(1)}/P^{(1)}$ and combining with the analogous limit for the bid side gives Eq. (5). $\square$

Proof of Proposition 3.1

As shown in Santos and Veronesi (2005), the pricing matrix B has the form

$$\frac{1}{d}\begin{bmatrix} L(1) + a(1-\bar{s}) & a(1-\bar{s}) \\ a\bar{s} & L(0) + a\bar{s} \end{bmatrix},$$

where L is given in the proposition and $d = L(0)L(1) + a\bar{s}L(1) + a(1-\bar{s})L(0)$. Since L and a are positive, so is d and hence all elements of B.

Starting from the pricing expressions shown in the text, Eq. (9) can be written

$$P_t^{(1)}\ X^{(1)} = C_t\ (B_{(1,0)}(1-s_t) + B_{(1,1)}s_t) \equiv C_t\, f(s_t)$$

and also as

$$P_t^{(1)} = D_t^{(1)}\ (B_{(1,0)}(1-s_t)/s_t + B_{(1,1)}) \equiv D_t^{(1)} g(s_t).$$

Defining the functions f and g is convenient notationally. Likewise, it is useful to define the total wealth to consumption ratio as a function of s:

$$P_t^{(0)} X^{(0)} + P_t^{(1)} X^{(1)} \equiv C_t\ F(s_t).$$

The computation of $\mathscr{I}$ is now straightforward. Suppressing the time subscript,

$$\begin{aligned} \mathscr{I} &= -\frac{X^{(1)}}{P^{(1)}}\frac{\mathrm{d}P^{(1)}}{\mathrm{d}X^{(1)}} \\ &= -\frac{X^{(1)}}{g(s)}\, g'(s)\, \frac{\mathrm{d}s}{\mathrm{d}X^{(1)}}. \end{aligned} \tag{A.1}$$

The derivative $\mathrm{d}s/\mathrm{d}X^{(1)}$ is to be evaluated under the value-neutral condition $\mathrm{d}X^{(0)}/\mathrm{d}X^{(1)} = -P^{(1)}/P^{(0)}$. Hence,

$$\begin{aligned} \frac{\mathrm{d}s}{\mathrm{d}X^{(1)}} &= \frac{D^{(1)}}{C_t} - \frac{s}{C_t}\frac{\mathrm{d}C_t}{\mathrm{d}X^{(1)}} \\ &= \frac{D^{(1)}}{C_t}\left(1 - \frac{X^{(1)}}{C_t}\left(D^{(1)} + D^{(0)}\frac{\mathrm{d}X^{(0)}}{\mathrm{d}X^{(1)}}\right)\right) \\ &= \frac{D^{(1)}}{C_t}\left(1 - s - \frac{X^{(1)}}{X^{(0)}}\frac{X^{(0)}D^{(0)}}{C_t}\frac{\mathrm{d}X^{(0)}}{\mathrm{d}X^{(1)}}\right) \\ &= \frac{D^{(1)}}{C_t}(1-s)\left(1 + \frac{P^{(1)}X^{(1)}}{P^{(0)}X^{(0)}}\right). \end{aligned}$$

This last expression may also be written

$$\frac{D^{(1)}}{C_t}(1-s)\left(\frac{F(s)}{F(s)-f(s)}\right).$$

Putting the pieces together,

$$\mathscr{I} = -s(1-s)\frac{g'(s)}{g(s)}\left(\frac{F(s)}{F(s)-f(s)}\right), \tag{A.2}$$

which holds generally for two-asset models in which s is the only state variable.

Using this and the explicit recipes for $g(.)$, $f(.)$, and $F(.)$ in the Santos and Veronesi (2005) case, $\mathscr{I}(s)$ becomes

$$\frac{(1-s)B_{1,0}\{(B_{0,0}+B_{1,0})(1-s)+(B_{0,1}+B_{1,1})s\}}{B_{0,0}B_{1,0}(1-s)^2+B_{0,1}B_{1,1}s^2+(B_{0,0}B_{1,1}+B_{0,1}B_{1,0})(1-s)s}.$$

The three statements in the proposition follow immediately from this formula. □

Proof of Proposition 3.2

Proposition 3 in Santos and Veronesi (2005) shows that, for each of the individual assets $i = 0, 1, \ldots, N$ as well as for the composite asset, M, consisting of a claim to assets 1 to N, the pricing equation in the log utility case is

$$\frac{P^{(i)}}{D^{(i)}} \equiv g^{(i)}(s^{(i)}) = k + \frac{k^{(i)}}{s^{(i)}},$$

where $k = (\phi + a)^{-1}$ and $k^{(i)} = ak\bar{s}^{(i)}/\phi$. Note that the ks are all positive.

As in the proof above, evaluating $\mathscr{I}^{(i)}$ requires computing $\mathrm{d}s^{(i)}/\mathrm{d}X^{(i)}$ subject to $\mathrm{d}X^{(0)}/\mathrm{d}X^{(i)} = -P^{(i)}/P^{(0)}$. The only difference with $N>1$ "risky" assets is that $s^{(0)} \neq (1-s^{(i)})$. So,

$$\begin{aligned}
\frac{\mathrm{d}s^{(i)}}{\mathrm{d}X^{(1)}} &= \frac{D^{(i)}}{C_t} - \frac{s^{(i)}}{C_t}\frac{\mathrm{d}C_t}{\mathrm{d}X^{(i)}} \\
&= \frac{D^{(i)}}{C_t}\left(1 - s^{(i)} - \frac{X^{(i)}}{X^{(0)}}\frac{X^{(0)}D^{(0)}}{C_t}\frac{\mathrm{d}X^{(0)}}{\mathrm{d}X^{(i)}}\right) \\
&= \frac{D^{(i)}}{C_t}(1-s^{(i)})\left(1 + \frac{s^{(0)}}{(1-s^{(i)})}\frac{P^{(i)}X^{(i)}}{P^{(0)}X^{(0)}}\right).
\end{aligned}$$

Now use Eq. (A.1), the fact that $P^{(i)}X^{(i)}/P^{(0)}X^{(0)} = s^{(i)}g^{(i)}/s^{(0)}g^{(0)}$, and the explicit forms of the price–dividend ratio to write

$$\begin{aligned}
\mathscr{I} &= -\frac{g^{(i)\prime}}{g^{(i)}}s^{(i)}(1-s^{(i)})\left(1 + \frac{s^{(0)}}{(1-s^{(i)})}\frac{s^{(i)}g^{(i)}}{s^{(0)}g^{(0)}}\right) \\
&= \frac{k^{(i)}}{ks^{(i)}+k^{(i)}}(1-s^{(i)})\left(1 + \frac{s^{(0)}}{(1-s^{(i)})}\frac{ks^{(i)}+k^{(i)}}{ks^{(0)}+k^{(0)}}\right).
\end{aligned} \tag{A.3}$$

If $i = M$, then $s^{(0)} = (1 - s^{(i)})$ and this is

$$\frac{k^{(M)}}{ks^{(M)} + k^{(M)}}(1 - s^{(M)})\left(1 + \frac{ks^{(M)} + k^{(M)}}{k(1 - s^{(M)}) + k^{(0)}}\right)$$

or

$$\frac{k^{(M)}(1 - s^{(M)})}{ks^{(M)} + k^{(M)}} + \frac{k^{(M)}(1 - s^{(M)})}{k(1 - s^{(M)}) + k^{(0)}}.$$

Clearly the first term is decreasing in $s^{(M)}$. The second term is as well because it is increasing in $s^{(0)}$. (For any positive constants, c_1, c_2, c_3, the function $c_1 x/(c_2 x + c_3)$ is increasing in x.) This shows that $((\mathrm{d}\mathscr{I}^{(M)})/\mathrm{d}s^{(M)})(s^{(M)}) < 0$.

Next, for $i \neq M$, consider the partial derivative of Eq. (A.3) with respect to $s^{(M)}$. In that case, $s^{(M)}$ appears only via the $s^{(0)} = (1 - s^{(M)})$ terms. Then the expression is just a positive constant times $s^{(0)}/ks^{(0)} + k^{(0)}$. Again, the derivative of this with respect to $s^{(0)}$ is always positive, so the derivative with respect to $s^{(M)}$ is negative, as claimed.

Finally, if $s^{(i)}$ is proportional to $s^{(M)}$ in Eq. (A.3) then we may put $s^{(0)} = (1 - \kappa s^{(i)})$ (where $\kappa > 0$ is the inverse of the λ in the proposition) and differentiate with respect to $s^{(i)}$, which clearly is equivalent to differentiating with respect to $s^{(M)}$. The expression to be differentiated is then

$$\frac{k^{(i)}}{ks^{(i)} + k^{(i)}}(1 - s^{(i)})\left(1 + \frac{(1 - \kappa s^{(i)})}{(1 - s^{(i)})}\frac{ks^{(i)} + k^{(i)}}{k(1 - s^{(i)}) + k^{(0)}}\right)$$

or

$$\frac{k^{(i)}(1 - s^{(i)})}{ks^{(i)} + k^{(i)}} + \frac{k^{(i)}(1 - \kappa s^{(i)})}{k(1 - \kappa s^{(i)}) + k^{(0)}}.$$

As above, both terms are decreasing in $s^{(i)}$, which establishes the third claim of the proposition.

Proof of Proposition 3.3

The limiting illiquidity can be computed by the following steps.

- Transform the independent variable in Eq. (12) to $w = 1/(1 + x)$ to examine the behavior at $z = 0$.
- Transform the dependent variable to $Q(w) = w\,P(w)$ and expand $Q(w)$ in powers of w in the neighborhood of $w = 0$, i.e.,

$$Q(w) = \sum_{n=0} q_n\, w^n.$$

- Plug this into the differential equation (12) and solve for the first two coefficients q_o and q_1 by setting the coefficient of each power to zero.

The last step also requires limiting values of $r(w)$ and $\dot{r(w)}$. So a similar expansion must be performed on Eq. (11) (after the same transformation of variables). I omit the tedious details.

The end result for the first two coefficients in the Q expansion is

$$q_0 = 1/(\phi - \mu(1-\gamma) + \tfrac{1}{2}\gamma(1-\gamma)\sigma^2)$$
$$q_1 = (1 + q_0(\mu - \gamma\sigma^2 - r_L(2 - \mu\gamma - \gamma\sigma^2 + \gamma(1+\gamma)\sigma^2)))/r_L,$$

where $r_L = \phi + \mu\gamma - \frac{1}{2}\gamma(1+\gamma)\sigma^2$.

In terms of $Q(w)$, the illiquidity $\mathscr{I}$ works out to be

$$[Q + (1-w)\dot{Q}]\left[\frac{w}{Q} + 1\right].$$

Hence, its limiting value is $Q(0) + \dot{Q}(0) = q_0 + q_1$. Simplifying the sum of expressions above leads to the value given in the proposition.

References

Acharya, V.V., Pedersen, L., 2004. Liquidity risk and asset pricing. Journal of Financial Economics 77, 375–410.

Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. Journal of Financial Markets 5 (1), 31–56.

Bagwell, L.S., 1992. Dutch auction repurchases: an analysis of shareholder heterogeneity. Journal of Finance 47 (1), 71–105.

Breen, W., Hodrick, L.S., Korajczyk, R., 2002. Predicting equity liquidity. Management Science 48, 470–483.

Campbell, J.Y., Grossman, S.J., Wang, J., 1993. Trading volume and serial correlation in stock returns. Quarterly Journal of Economics 108, 905–939.

Chordia, T., Roll, R., Subrahmanyam, A., 2000. Commonality in liquidity. Journal of Financial Economics 56, 3–28.

Chordia, T., Sarkar, A., Subrahmanyam, A., 2004. An empirical analysis of stock and bond market liquidity. Review of Financial Studies 18, 85–129.

Cochrane, J., Longstaff, F., Santa Clara, P., 2004. Two trees: asset price dynamics induced by market clearing. Unpublished working paper, University of Chicago.

Eckbo, B.E., Norli, O., 2002. Pervasive liquidity risk. Unpublished working paper, Tuck School, Dartmouth.

Eisfeldt, A., 2004. Endogenous liquidity in asset markets. Journal of Finance 59, 1–30.

Fleming, M.J., 2001. Measuring treasury market liquidity. Federal Reserve Bank of New York, Staff Paper 133.

Fujimoto, A., 2004. Macroeconmic sources of systematic liquidity. Unpublished working paper, University of Alberta.

Garman, M.B., 1976. Market microstructure. Journal of Financial Economics 3, 257–275.

Glosten, L.R., Harris, L.E., 1988. Estimating the components of the bid/ask spread. Journal of Financial Economics 21, 123–142.

Hasbrouck, J., Seppi, D., 2001. Common factors in prices order flows and liquidity. Journal of Financial Economics 59, 383–411.

Huberman, G., Halka, D., 2001. Systematic liquidity. Journal of Financial Research 24, 161–178.

Jones, C.M., 2002. A century of stock market liquidity and trading costs. Unpublished working paper, Columbia University.

Loderer, C., Cooney, J.W., VanDrunen, L.D., 1991. The price elasticity of demand for common stock. Journal of Finance 46 (2), 621–651.

Lucas, R., 1978. Asset prices in an exchange economy. Econometrica 46, 1429–1445.

Menzly, L., Santos, T., Veronesi, P., 2004. Understanding predictability. Journal of Political Economy 112, 1–47.

Novy-Marx, R., 2004. On the excess returns to illiquidity. Unpublished working paper, University of Chicago.
Pástor, L., Stambaugh, R., 2003. Liquidity risk and expected stock returns. Journal of Political Economy 111, 642–685.
Santos, T., Veronesi, P., 2005. Labor income and predictable stock returns. Review of Financial Studies, Forthcoming.
Shleifer, A., 1986. Do demand curves for stocks slope down? Journal of Finance 41, 579–590.
Stoll, H.R., 1978. The supply of dealer services in security markets. Journal of Finance 33, 1133–1151.