

5-10-2015

# A Search Model of the Aggregate Demand for Safe and Liquid Assets

Ji Shen

*London School of Economics & Political Science*

Hongjun Yan

*DePaul University*

---

## Recommended Citation

Shen, Ji and Yan, Hongjun, A Search Model of the Aggregate Demand for Safe and Liquid Assets (May 10, 2015). Available at SSRN: <https://ssrn.com/abstract=2384895> or <http://dx.doi.org/10.2139/ssrn.2384895>

This Article is brought to you for free and open access by the Driehaus College of Business at Via Sapientiae. It has been accepted for inclusion in Publications – Dreihaus College of Business by an authorized administrator of Via Sapientiae. For more information, please contact [wsulliv6@depaul.edu](mailto:wsulliv6@depaul.edu), [c.mcclure@depaul.edu](mailto:c.mcclure@depaul.edu).

# A Search Model of the Aggregate Demand for Safe and Liquid Assets\*

Ji Shen  
London School of Economics

Hongjun Yan  
Yale School of Management

May 10, 2015

\*We thank Nicolae Garleanu, Arvind Krishnamurthy, Andrew Metrick, Artem Neklyudov, Dimitri Vayanos, Pierre-Olivier Weill, Randall Wright, and seminar participants at CKGSB, Yale, EFA, and European Summer Symposium in Financial Markets for helpful comments. Please direct all correspondence to Hongjun Yan, Email: [hongjun.yan@yale.edu](mailto:hongjun.yan@yale.edu). The latest version of the paper is available at <http://faculty.som.yale.edu/hongjunyan/>.

# A Search Model of the Aggregate Demand for Safe and Liquid Assets

## Abstract

Safe and liquid assets, such as Treasury bonds, are money-like instruments that command a convenience yield. We analyze this in a search model of two assets that differ in liquidity and safety. In contrast to the reduced-form approach, which puts the safe and liquid asset in utility function, we explicitly model investors' trading needs and the trading friction. One new implication from this approach is that the marginal investor's preference for safety and liquidity is not enough in determining the premium. Instead, the distribution of investors' preferences plays a direct role. Our model implies that an increase in the supply of the liquid asset may increase or decrease the liquidity premium, depending on the distribution of investors' liquidity preference. Our model shows that investors may over- or underinvest in the search technology relative to a central planner, and that overinvestment occurs when investors' expected trading frequency is in the intermediate region.

*JEL* Classification Numbers: G11, G23.

*Keywords:* Convenience Yield, Safe and Liquid Asset, Search.

# 1 Introduction

There has been growing interest in the role of “safe and liquid assets” in a financial system, especially since the recent financial crisis. One finding that emerges from these studies is that safe and liquid assets, such as Treasury bonds, are like money, commanding a sizeable premium for their safety and liquidity (Krishnamurthy and Vissing-Jorgensen 2012). What are the determinants of this premium? How does the supply of Treasury bonds affect the premium? When risky assets become more liquid, how does it affect their own prices, as well as the Treasury price? What is the welfare implication when traders invest to improve the liquidity of risky assets?

One framework for addressing these questions is a representative agent model. For example, Krishnamurthy and Vissing-Jorgensen (2012) follow the tradition of money-in-the-utility-function formulation (e.g., Sidrauski 1967) and include the Treasury holding in the representative investor’s utility function. In equilibrium, the liquidity premium is determined such that the representative agent is indifferent between holding the Treasury and a less liquid asset. That is, the representative agent is the marginal investor whose indifference condition determines the liquidity premium. The appeal of this approach is its simplicity, and one can analyze the liquidity premium without explicitly modeling investors’ trading needs and trading frictions.

We adopt an alternative framework, and explicitly model investors’ trading needs and trading frictions. Not only does this make it possible to directly connect liquidity premium to trading frictions—it also leads to new implications that are absent in the representative agent framework. Specifically, the marginal investor’s liquidity preference is no longer enough to determine the premium. Instead, the distribution of investors’ liquidity preferences also plays a direct role. For example, we find that an increase in the supply of Treasury bonds may increase or decrease their liquidity premium, depending on the distribution of investors’ liquidity preferences.

The intuition is as follows. Suppose assets 1 and 2 have identical cash flows, but asset 2 is “more liquid” than asset 1. In the reduced-form approach, asset 2 being more liquid is modeled as investors deriving a “convenience yield” from holding asset 2 (i.e., putting the holding of asset 2 in an investor’s utility function). Let  $P_1$  and  $P_2$  be the prices of assets 1 and 2, respectively. The liq-

uidity premium,  $P_2 - P_1$ , is determined by the present value of the marginal investor's convenience yield. Hence, the marginal investor's liquidity preference fully determines the premium.

However, this is no longer the case once we explicitly take trading frictions into account. Suppose that asset 2 is perfectly liquid, and that the friction for trading asset 1 is that investors need to search in the market and can trade only when they meet their counterparties. In this case, the marginal investor's liquidity preference cannot fully determine the premium. To see this, suppose that  $P_1$  decreases by one dollar due to a reduction of demand from its investors. We will see that, if the marginal investor between assets 1 and 2 remains the same,  $P_2$  will decrease by *less* than one dollar, and hence the liquidity premium  $P_2 - P_1$  will increase. The reason is that the marginal investor's value function is *less* sensitive to  $P_1$  than to  $P_2$ : Intuitively, since asset 2 is perfectly liquid,  $P_2$  is the price at which an investor can transact right away. So, a one-dollar drop in  $P_2$  leads to a one-dollar increase in his value function. In contrast, a one-dollar drop in  $P_1$  leads to a *less-than-one-dollar* increase in his value function. This is due to the trading friction:  $P_1$  is the price at which the investor can transact only when he meets his counterparty. There is a chance that the investor cannot find his counterparty before his trading need disappears. This point arises naturally once we explicitly account for the trading friction, but is absent in the reduced-form approach that abstracts away from trading frictions.

In essence, the notion of "market price" is different in a setup where frictions are modeled explicitly than in a setup that treats frictions implicitly. In a model which treats frictions only implicitly, the market price is the price at which investors can transact at immediately. However, this is not the case in models with explicit trading frictions.

We formalize the above intuition by extending the over-the-counter (OTC) market model of Duffie, Garleanu, and Pedersen (2005) by introducing two assets. In the baseline model, the two assets are claims to identical cash flows but have different liquidity. Asset 1 (e.g., agency debt) is less liquid, and trade occurs only when a buyer meets a seller. In contrast, asset 2 (e.g., Treasury) is perfectly liquid and transactions occur without any delay. There is a continuum of investors, whose trading needs are due to the changes of their valuations of the two assets. In particular, when a type- $\Delta$  investor receives \$1 from asset 1 or 2, he derives a utility of  $1 + \Delta$ . We normalize

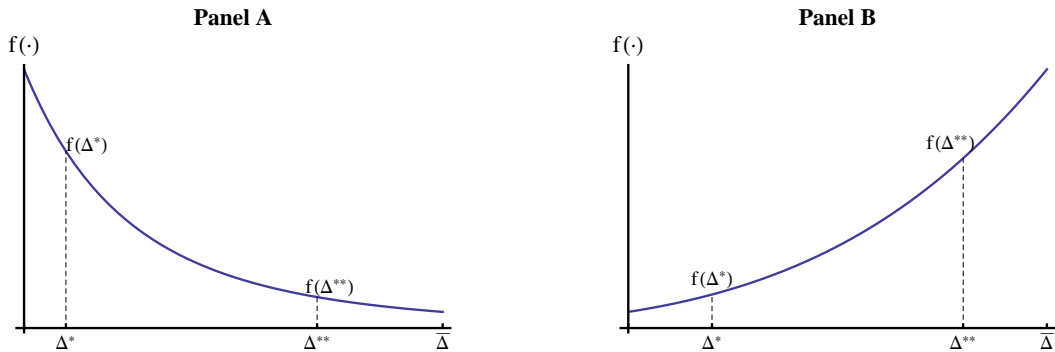
the region for investors' possible types to  $[0, \bar{\Delta}]$ . An investor's type stays constant until the arrival of a shock. Once the shock arrives, his new type is drawn from a random variable, which has a density function of  $f(\cdot)$  on  $[0, \bar{\Delta}]$ . Investors' types are independent from one another. Hence, in the steady state,  $f(\cdot)$  is also the cross-sectional distribution of investors' types.

We show that, in equilibrium, there are two cutoff points,  $\Delta^*$  and  $\Delta^{**}$ , with  $0 < \Delta^* < \Delta^{**} < \bar{\Delta}$ . Investors with high types (i.e.,  $\Delta \in (\Delta^{**}, \bar{\Delta}]$ ) choose to buy asset 2, those with intermediate types (i.e.,  $\Delta \in (\Delta^*, \Delta^{**})$ ) choose to buy asset 1, and those with low types (i.e.,  $\Delta \in [0, \Delta^*)$ ) choose not to buy any asset. Investors  $\Delta^*$  and  $\Delta^{**}$  are marginal investors: investor- $\Delta^{**}$  is indifferent between buying asset 1 and buying asset 2, while investor- $\Delta^*$  is indifferent between buying asset 1 and not buying any asset.

The liquidity preference of the marginal investor between the two assets (i.e.,  $\Delta^{**}$ ) affects the liquidity premium, but, as explained earlier, it cannot fully pin down the liquidity premium. We find that the liquidity premium increases in  $\Delta^{**}$  but decreases in  $\Delta^*$ . Intuitively, a higher  $\Delta^{**}$  means that trading delay is more costly for the investor. Hence, asset 2 commands a higher premium. How does  $\Delta^*$  affect the liquidity premium? Since investor- $\Delta^*$  is the marginal investor between investing asset 1 and not investing, holding everything else constant, a decrease in  $\Delta^*$  decreases  $P_1$ . In response to this drop in  $P_1$ , as noted earlier,  $P_2$  would decrease less than  $P_1$  does. That is, the liquidity premium  $P_2 - P_1$  increases when  $\Delta^*$  decreases.

Our model implies that an increase in the supply of asset 2 may increase or decrease the liquidity premium, depending on the distribution  $f(\cdot)$ . Intuitively, when the supply of asset 2 increases, it attracts more investors with high  $\Delta$ , pushing down both  $\Delta^{**}$  and  $\Delta^*$ . As noted earlier, the liquidity premium increases in  $\Delta^{**}$  but decreases in  $\Delta^*$ . In the case illustrated in Panel A of Figure 1, for example,  $f(\Delta^*)$  is significantly larger than  $f(\Delta^{**})$ . That is, there are many investors whose  $\Delta$  is around  $\Delta^*$ , but very few investors around  $\Delta^{**}$ . When the supply of asset 2 increases,  $\Delta^{**}$  decreases significantly, but  $\Delta^*$  decreases only slightly. Hence, the impact from  $\Delta^{**}$  dominates, and the increase in the supply of asset 2 decreases the liquidity premium. Similarly, in the case illustrated in Panel B of Figure 1,  $f(\Delta^*)$  is significantly lower than  $f(\Delta^{**})$ . The impact from  $\Delta^*$  dominates, and the increase in the supply of asset 2 increases the liquidity

premium.



**Figure 1:** Distribution of liquidity preferences across investors  $f(\cdot)$ .

What are the empirical implications from this result? Suppose we interpret asset 2 as Treasury bonds and asset 1 as agency bonds or highly rated corporate bonds. Then, it might be reasonable to think this case is summarized by Panel A: a small fraction of investors have very high  $\Delta$ . For example, commercial banks can use Treasury securities as collateral to issue checking accounts, and hedge funds can use them as collateral for their derivative positions. For most investors, however, their  $\Delta$  is modest. In this case, the increase in Treasury supply decreases the yield spreads between Treasury and highly rated bonds, as documented in Krishnamurthy and Vissing-Jorgensen (2012). On the other hand, if we interpret asset 1 as junk bonds and asset 2 as bonds with investment-grades and above (e.g., investment-grade rated corporate bonds, agency bonds and Treasury securities), the case is more likely to correspond to Panel B, where very few specialized investors (such as hedge funds) are the marginal investors for asset 1 (i.e.,  $f(\Delta^*)$  is small). With this interpretation, our model implies that the increase of the supply of bonds with investment-grades and above increases the spread between junk bonds and investment-grade bonds.

When the search friction in market 1 is alleviated, how does it affect  $P_1$  and  $P_2$ ? Our model shows that it decreases  $P_2$ , because when trading asset 1 is easier, asset 2 becomes relatively less appealing. Moreover, the liquidity improvement in market 1 has a mixed effect on the price of asset 1. Intuitively, when search becomes slower, sellers in market 1 are willing to accept a lower

price to speed up their transactions. Similarly, buyers are willing to offer a higher price to reduce their waiting time. Hence, the total impact is mixed, and depends on which side is more eager to speed up the transaction.

Our welfare analysis on the investment in the search technology for market 1 shows that investors may over- or underinvest relative to a central planner. The reason is that the investment has two externalities. First, when an investor improves his search technology, it not only benefits himself, but also benefits his potential trading partners. This leads to a free-riding problem and underinvestment. Second, investment in the search technology helps more investors to execute their trades, and so reduces the number of investors in the market, making it more difficult for all investors to meet their counterparties. Investors don't internalize this negative externality and so overinvest relative to a central planner. Hence, the tradeoff between the two effects determines whether investors over- or underinvest in their search technology. We find that overinvestment tends to occur when investors' expected trading frequency is in the intermediate region.

## 1.1 Related Literature

Our paper belongs to the recent literature that analyzes OTC markets in the search framework developed by Duffie, Garleanu, and Pedersen (2005). This framework has been extended to include risk-averse agents (Duffie, Garleanu, and Pedersen (2007)), unrestricted asset holdings (Lagos and Rocheteau (2009)). It has also been adopted to analyze a number of issues, such as security lending (Duffie, Garleanu, and Pedersen (2002)), liquidity provision (Weill (2007)), on-the-run premium (Vayanos and Wang (2007), Vayanos and Weill (2008)), cross-sectional returns (Weill (2008)), portfolio choices (Garleanu (2009)), liquidity during a financial crisis (Lagos, Rocheteau, and Weill (2011)), price pressure (Feldhutter (2012)), order flows in an OTC market (Lester, Rocheteau, and Weill, (2014)), commercial aircraft leasing (Gavazza 2011), high frequency trading (Pagnotta and Philippon (2013)), the roles of benchmarks in OTC markets (Duffie, Dworczak, and Zhu (2014)), adverse selection and repeated contacts in opaque OTC markets (Zhu (2012)), intermediation chains (Afonso and Lagos (2015), Hugonnier, Lester, and Weill (2014), Shen, Wei, and Yan (2015)), trading network structure (Neklyudov (2014)), as well as the interaction



between corporate default decision and liquidity (He and Milbradt (2013)). Another literature follows Kiyotaki and Wright (1993) to analyze the liquidity value of money. In particular, Lagos and Wright (2005) develop a tractable framework that has been adopted to analyze liquidity and asset pricing (e.g., Lagos (2010), Lester, Postlewaite, and Wright (2012), and Li, Rocheteau, and Weill (2012), Lagos and Zhang (2014)). Trejos and Wright (2014) synthesize this literature with the studies under the framework of Duffie, Garleanu, and Pedersen (2005).

Our paper is related to these studies, and one distinctive feature is our analysis of the supply effect on the premium. Another insight from our model is the contrast between the reduced-form approach and the search approach that explicitly accounts for trading frictions. This is parallel to the point stressed in the classical search-theoretical model of Kiyotaki and Wright (1989), which emphasizes the importance of explicitly modeling the frictions that render money essential. This idea has led to the so-called New Monetarist Economics, which emphasize that assets are valued not only for their fundamentals (i.e., claims to consumption goods) but also for their liquidity—the extent to which they facilitate exchange in an imperfect market (see Williamson and Wright (2010, 2011) for recent surveys).

## 2 The Model

Time is continuous and goes from 0 to  $\infty$ . There is a continuum of investors, and the total population size is  $N$ . They have access to a riskless bank account with an interest rate  $r$ . There are two assets, assets 1 and 2, which are traded in two separate markets. The supplies for assets 1 and 2 are  $X_1$  and  $X_2$ , respectively, and  $X_1 + X_2 < N$ . The two assets have the same cash flows, and each unit of the asset pays \$1 per unit of time until infinity. However, asset 1 is less liquid than asset 2.

Our formulation of the market for asset 1 follows Garleanu (2009) and Lagos and Rocheteau (2009). In this market, investors face a potential delay in finding market makers. Once they meet a market maker, they can execute their trades and take the price  $P_1$  as given. The potential delay is as follows. Let  $\mu_1^b$  and  $\mu_1^s$  be the measures of buyers and sellers in the market for asset 1, and both will be determined endogenously in equilibrium. A buyer meets a market maker at the rate

$\lambda\mu_1^s$ , where  $\lambda > 0$  is a constant. That is, during  $[t, t + dt)$  a buyer meets a market maker with a probability  $\lambda\mu_1^s dt$ . Similarly, a seller meets a market maker who can buy his asset at the rate  $\lambda\mu_1^b$ . Hence, the total number of trades per unit of time is  $\lambda\mu_1^s\mu_1^b$ . The search friction reduces when  $\lambda$  increases, and completely disappears when  $\lambda$  goes to infinity.

This formulation is a slight modification of that in Garleanu (2009) and Lagos and Rocheteau (2009). Specifically, we assume that the arrival rate of the market maker depends on the population size of the investors on the other side of the market. For example, for a buyer, the larger the seller population  $\mu_1^s$ , the quicker the buyer is expected to find a market maker to sell him the asset. This captures the notion that an investor faces a shorter delay if there are more investors trying to be on the other side of the transaction.<sup>1</sup>

The market for asset 2 is more liquid. To simplify our analysis, we let the search technology in market 2 go to perfection, i.e., investors in market 2 can trade instantly.<sup>2</sup>

## 2.1 Trading needs

Investors have different types, and their types may change over time. If an investor's current type is  $\Delta$ , he derives a utility  $1 + \Delta$  when receiving the \$1 coupon from either asset. One interpretation for a positive  $\Delta$  is that some investors, such as insurance companies, have a strong preference for long-term bonds, as modeled in Vayanos and Vila (2009). Another interpretation is that some investors can benefit from using those assets as collateral and so value them more, as discussed in Bansal and Coleman (1996) and Gorton (2010). An interpretation of a negative  $\Delta$  can be that the investor suffers a liquidity shock and so finds it costly to carry the asset on his balance sheet. We assume that  $\Delta$  can take any value in a closed interval. Without loss of generality, we can normalize the interval to  $[0, \bar{\Delta}]$ .

Each investor's type changes independently with intensity  $\kappa$ . That is, during  $[t, t + dt)$ , with a probability  $\kappa dt$ , an investor's type changes and is independently drawn from a random variable, which has a probability density function  $f(\cdot)$  on the support  $[0, \bar{\Delta}]$ , with  $f(\Delta) < \infty$  for any

---

<sup>1</sup>We also solve our model without this modification. All our main results, except for the welfare implication in Section 2.8, remain similar.

<sup>2</sup>We also solved a version of the model in which the search technology in market 2 is imperfect but is better than the one in market 1. All our results remain similar.

$\Delta \in [0, \bar{\Delta}]$ . We use  $F(\cdot)$  to denote the corresponding cumulative distribution function.

The changes in investors' types make them trade the two assets. Following Duffie, Garleanu, and Pedersen (2005) and Vayanos and Wang (2007), we assume each investor can hold either 0 or 1 unit of only one of the assets.<sup>3</sup> Hence, an investor can buy an asset only when he currently does not hold either asset, and can sell an asset only if he is currently holding the asset. All investors are risk-neutral and share the same time discount rate  $r$ . An investor's objective function is given by

$$\sup_{\theta_{1\tau}, \theta_{2\tau}} \mathbf{E}_t \left[ \int_t^\infty e^{-r(\tau-t)} ((\theta_{1\tau} + \theta_{2\tau})(1 + \Delta_\tau) d\tau - P_{1\tau} d\theta_{1\tau} - P_{2\tau} d\theta_{2\tau}) \right],$$

where  $\theta_{1\tau}$  and  $\theta_{2\tau}$  are the investor's holdings in assets 1 and 2 at time  $\tau$ ;  $\Delta_\tau$  is the investor's type at time  $\tau$ ; and  $P_{i\tau}$ , for  $i = 1, 2$ , is asset  $i$ 's price at time  $\tau$  and will be determined in equilibrium.

## 2.2 Demographics

Investors can be classified into three categories: owners of asset 1 ( $\theta_{1t} = 1$  and  $\theta_{2t} = 0$ ), owners of asset 2 ( $\theta_{1t} = 0$  and  $\theta_{2t} = 1$ ), and non-owners (i.e.,  $\theta_{1t} = \theta_{2t} = 0$ ). This section describes each category in detail.

A non-owner with a type  $\Delta$  has three choices: search to buy asset 1, buy asset 2, or stay inactive. We conjecture and verify later that a non-owner's optimal choice can be summarized as

$$\begin{cases} \text{stay inactive if } \Delta \in [0, \Delta_0^*), \\ \text{search to buy asset 1 if } \Delta \in (\Delta_0^*, \Delta_0^{**}), \\ \text{buy asset 2 if } \Delta \in (\Delta_0^{**}, \bar{\Delta}]. \end{cases} \quad (1)$$

That is, he buys asset 2 if  $\Delta > \Delta_0^{**}$ , searches to buy asset 1 if  $\Delta \in (\Delta_0^*, \Delta_0^{**})$ , and stays inactive if  $\Delta < \Delta_0^*$ . A non-owner is indifferent between staying inactive and searching to buy asset 1 at  $\Delta_0^*$ , and is indifferent between searching to buy asset 1 and buying asset 2 at  $\Delta_0^{**}$ . Note that due to the search friction in market 1, the buyers of asset 1 face a delay in their transactions. In the meantime, their types may change, and then they will adjust their actions accordingly. In market 2, however, the buyers become owners of asset 2 instantly.

---

<sup>3</sup>This deviates from the formulation in Garleanu (2009) and Lagos and Rocheteau (2009), where the asset holdings are not restricted. We keep this traditional assumption on asset holdings for tractability. We impose the same asset holding restriction in both markets to isolate the effects from the search friction in market 1. More generally, in the case where the search technology in market 2 is imperfect, this formulation isolates the effects from the difference in the search frictions across the two markets.

An owner of asset 1 has two choices: search to sell asset 1 or hold on to it. We conjecture and later verify that this investor's optimal choice can be summarized as

$$\begin{cases} \text{search to sell his asset if } \Delta \in [0, \Delta_1^*), \\ \text{hold on to his asset if } \Delta \in (\Delta_1^*, \bar{\Delta}]. \end{cases} \quad (2)$$

That is, he searches to sell asset 1 if  $\Delta < \Delta_1^*$ , holds on to the asset if  $\Delta > \Delta_1^*$ , and is indifferent between the two actions if his type is  $\Delta_1^*$ . Moreover, investors face a delay in selling their asset 1. In the meantime, their types may change, and they may need to adjust their actions accordingly. If an investor succeeds in selling his asset 1, he becomes a non-owner and faces the three choices described in equation (1).

An owner of asset 2 also has two choices: sell it or hold on to it. We conjecture and later verify that this investor's optimal choice can be summarized as

$$\begin{cases} \text{sell his asset if } \Delta \in [0, \Delta_2^*), \\ \text{hold on to his asset if } \Delta \in (\Delta_2^*, \bar{\Delta}]. \end{cases} \quad (3)$$

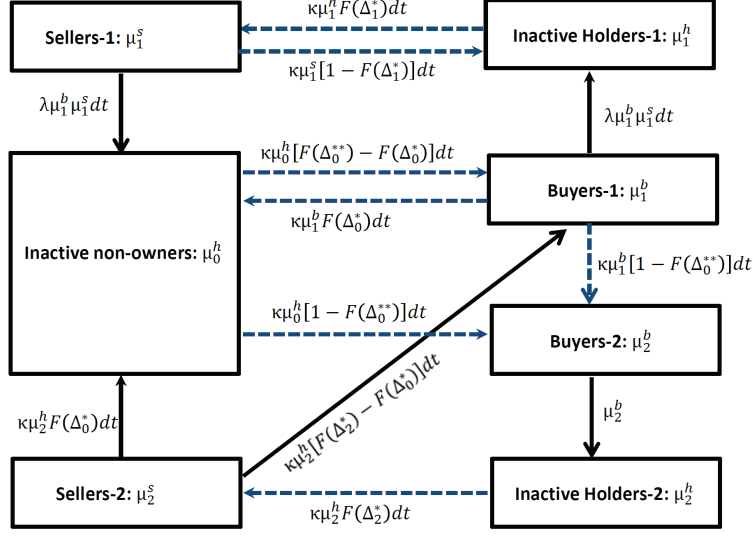
That is, he sells asset 2 if  $\Delta < \Delta_2^*$ , holds on to the asset if  $\Delta > \Delta_2^*$ , and is indifferent between the two actions if his type is  $\Delta_2^*$ . Since there is no search friction in market 2, investors can execute their transactions right away.

Due to the change in  $\Delta$  and execution of his trade, an investor's status changes over time. We now describe the evolution of the population sizes of each category of investors. Since we will focus on the steady-state equilibrium, we will omit the time subscript for the population size of each group of investors. For  $i = 1, 2$ , we use  $\mu_i^s$  to denote the population size of the sellers for asset  $i$ , and use  $\mu_i^b$  to denote the population size of the buyers for asset  $i$ . Similarly, we use  $\mu_i^h$ , for  $i = 0, 1, 2$ , to denote the population sizes of the inactive investors who are non-owners, owners of asset 1, and owners of asset 2, respectively. Hence, there are seven groups of investors.

Figure 2 illustrates investors' migration across the seven groups. For sellers of asset 1, for example, the inflow to this group during the period  $[t, t + dt)$  is  $\mu_1^h \kappa F(\Delta_1^*) dt$ , since  $\kappa F(\Delta_1^*)$  is the intensity for an inactive asset 1 holder to become a seller (i.e., his type becomes lower than  $\Delta_1^*$ ). The outflow from the group of asset-1 sellers has two components. First, during the period  $[t, t + dt)$ ,  $\lambda \mu_1^b \mu_1^s dt$  investors succeed in selling their asset 1 and become inactive non-owners. Second,  $\kappa \mu_1^s [1 - F(\Delta_1^*)] dt$  investors do not want to sell asset 1 any more because their types now

become higher than  $\Delta_1^*$ . In the steady state, the inflow equals the outflow:

$$\mu_1^h \kappa F(\Delta_1^*) = \lambda \mu_1^b \mu_1^s + \kappa \mu_1^s [1 - F(\Delta_1^*)]. \quad (4)$$



**Figure 2:** This plot illustrates each investor group's size and inflows and outflows. The black solid arrows denote the flows induced by trading, and the blue dash arrows denote the flows due to the changes in investors' types.

Applying the same logic to the buyers of asset 1, inactive owners of asset 1, and inactive non-owners, we obtain the following:

$$\kappa \mu_0^h [F(\Delta_0^{**}) - F(\Delta_0^*)] + \kappa \mu_2^h [F(\Delta_2^*) - F(\Delta_0^*)] = \lambda \mu_1^b \mu_1^s + \kappa \mu_1^b [F(\Delta_0^*) + 1 - F(\Delta_0^{**})], \quad (5)$$

$$\kappa \mu_1^s [1 - F(\Delta_1^*)] + \lambda \mu_1^b \mu_1^s = \kappa \mu_1^h F(\Delta_1^*), \quad (6)$$

$$\lambda \mu_1^b \mu_1^s + \kappa (\mu_1^b + \mu_2^h) F(\Delta_0^*) = \kappa \mu_0^h [1 - F(\Delta_0^*)]. \quad (7)$$

Following Garleanu (2009) and Lagos and Rocheteau (2009), we also assume that the market makers do not hold inventory and simply serve as match makers. This implies that

$$\mu_1^b = \mu_1^s. \quad (8)$$

Market 2 has no search friction, the measures of buyers and sellers are infinitesimal,

$$\mu_2^b = \kappa (\mu_0^h + \mu_1^b) [1 - F(\Delta_0^{**})] dt \quad (9)$$

$$\mu_2^s = \kappa \mu_2^h F(\Delta_2^*) dt, \quad (10)$$

and during each instant  $[t, t + dt)$ , the flow of buyers is equal to the flow of sellers

$$\left(\mu_0^h + \mu_1^b\right) [1 - F(\Delta_0^{**})] = \mu_2^h F(\Delta_2^*). \quad (11)$$

Finally, the investors in all groups add up to the total population:

$$\mu_1^h + \mu_1^s + \mu_1^b + \mu_2^h + \mu_2^s + \mu_2^b + \mu_0^h = N. \quad (12)$$

### 2.3 Value functions

For the case  $\theta_{1t} = \theta_{2t} = 0$  (i.e., the investor is a non-owner), we use  $V_1^b(\Delta)$ ,  $V_2^b(\Delta)$ , and  $V_0^h(\Delta)$  to denote the investor's expected utility if he chooses to buy asset 2, to search to buy asset 1, and to stay inactive, respectively. For the case  $\theta_{1t} = 1$  and  $\theta_{2t} = 0$  (i.e., the investor is an owner of asset 1), we use  $V_1^s(\Delta)$  and  $V_1^h(\Delta)$  to denote the investor's expected utility if he searches to sell asset 1, and to keep asset 1, respectively. For the case  $\theta_{1t} = 0$  and  $\theta_{2t} = 1$  (i.e., the investor is an owner of asset 2), we use  $V_2^s(\Delta)$  and  $V_2^h(\Delta)$  to denote the investor's expected utility if he chooses to sell asset 2, and to keep asset 2, respectively. In the steady state, these expected utilities are time-invariant, implying the following:

$$V_1^b(\Delta) = \frac{\lambda\mu_1^s [V_1^h(\Delta) - P_1] + \kappa\mathbf{E} [\max \{V_1^b(\Delta'), V_2^b(\Delta'), V_0^h(\Delta')\}]}{\lambda\mu_1^s + \kappa + r}, \quad (13)$$

$$V_1^h(\Delta) = \frac{1 + \Delta + \kappa\mathbf{E} [\max \{V_1^s(\Delta'), V_1^h(\Delta')\}]}{\kappa + r}, \quad (14)$$

$$V_1^s(\Delta) = \frac{1 + \Delta + \lambda\mu_1^b \max\{V_0^h(\Delta), V_2^b(\Delta)\} + \lambda\mu_1^b P_1 + \kappa\mathbf{E}[\max\{V_1^s(\Delta'), V_1^h(\Delta')\}]}{\lambda\mu_1^b + \kappa + r}, \quad (15)$$

$$V_2^b(\Delta) = V_2^h(\Delta) - P_2, \quad (16)$$

$$V_2^s(\Delta) = \max \{V_0^h(\Delta), V_1^b(\Delta)\} + P_2, \quad (17)$$

$$V_2^h(\Delta) = \frac{1 + \Delta + \kappa\mathbf{E} [\max \{V_2^s(\Delta'), V_2^h(\Delta')\}]}{\kappa + r}, \quad (18)$$

$$V_0^h(\Delta) = \frac{\kappa}{\kappa + r} \mathbf{E} [\max \{V_1^b(\Delta'), V_2^b(\Delta'), V_0^h(\Delta')\}]. \quad (19)$$

### 2.4 Prices with trading fictions

Once we explicitly account for the trading friction, the notion of the price of an asset is different that in a reduced-form model. For example, an holder of asset 1 can no longer exchange the asset for  $P_1$  instantly. This straight forward but easy-to-overlook feature implies that investors' value

functions have different sensitivities to  $P_1$  and  $P_2$ . From equation (13), we obtain the following lemma.

**Lemma 1** *An investor's expected utility is more sensitive to  $P_2$  than to  $P_1$ :  $\frac{\partial V_2^b(\Delta)}{\partial P_2} = -1$  and  $\frac{\partial V_1^b(\Delta)}{\partial P_1} = -\frac{\lambda\mu_1^s}{\lambda\mu_1^s + \kappa + r}$ .*

The intuition is the following. The market for asset 2 is perfectly liquid, i.e., a buyer can pay  $P_2$  to get asset 2 right away. Hence, holding everything else constant, a one-dollar drop in  $P_2$  increases the investor's expected utility by one dollar. In contrast, a one-dollar drop in  $P_1$  does not mean the investor gets a one-dollar benefit. This is because the market for asset 1 is illiquid, and the investor may not be able to benefit fully from the price drop. Due to the delay in searching, the investor can only enjoy the benefit in the future. Moreover, the investor may not be able to benefit at all if he cannot meet a seller before his  $\Delta$  changes and his demand disappears. As a result, the investor's expected utility is less sensitive to  $P_1$ .

This intuition is absent in the money-in-the-utility-function formulation, where the trading friction is not explicitly modeled and the notion of liquidity is captured by putting the liquid asset directly into investors' utility function. Hence, the sensitivity of the buyer's expected utility to price is still one-to-one: a one-dollar drop in price increases the expected utility by one dollar. The essence is that the notion of market price is different in a setup where frictions are modeled explicitly than in a setup that treats frictions implicitly. In models with explicit trading frictions, the market price is not the price at which investors can transact at immediately.

## 2.5 Equilibrium

**Definition 1** *A steady-state equilibrium consists of asset prices  $P_1$  and  $P_2$ , the cutoff points  $(\Delta_0^*, \Delta_0^{**}, \Delta_1^*, \Delta_2^*)$ , such that*

- 1) *the sizes of each group  $(\mu_1^h, \mu_1^s, \mu_1^b, \mu_2^h, \mu_2^s, \mu_2^b, \mu_0^h)$  remain constants over time, i.e., satisfy (4)–(12);*
- 2) *the choices implied by (1)–(3) and (13)–(19) are optimal for all investors;*

3) both markets clear:

$$X_1 = \mu_1^h + \mu_1^s. \quad (20)$$

$$X_2 = \mu_2^h. \quad (21)$$

**Proposition 1** *The steady-state equilibrium for the above economy is the following. The cutoff points are given by*

$$\Delta_0^* = \Delta_1^* = \Delta^*,$$

$$\Delta_2^* = \Delta_0^{**} = \Delta^{**},$$

where

$$\Delta^* = F^{-1} \left( 1 - \frac{X_1 + X_2}{N} \right), \quad (22)$$

$$\Delta^{**} = F^{-1} \left( 1 - \frac{X_2}{N - X_1} \right). \quad (23)$$

The population sizes for each group are given by

$$\mu_1^s = \mu_1^b = \mu_1, \quad (24)$$

$$\mu_1^h = X_1 - \mu_1, \quad (25)$$

$$\mu_0^h = N - X_2 - X_1 - \mu_1, \quad (26)$$

$$\mu_2^h = X_2, \quad (27)$$

$$\mu_2^b = \kappa X_2 \left( 1 - \frac{X_2}{N - X_1} \right) dt \quad (28)$$

$$\mu_2^s = \kappa X_2 \left( 1 - \frac{X_2}{N - X_1} \right) dt, \quad (29)$$

where

$$\mu_1 \equiv \sqrt{\left( \frac{\kappa}{2\lambda} \right)^2 + \frac{\kappa X_1}{\lambda} \left( 1 - \frac{X_1 + X_2}{N} \right)} - \frac{\kappa}{2\lambda}. \quad (30)$$

The asset prices are given by

$$P_1 = \frac{1 + \Delta^*}{r} + \frac{\kappa \int_{\Delta^*}^{\Delta^{**}} [1 - F(\Delta)] d\Delta}{r (\lambda \mu_1 + \kappa + r)} - \frac{\kappa \int_0^{\Delta^*} F(\Delta) d\Delta}{r (\lambda \mu_1 + \kappa + r)}, \quad (31)$$

$$P_2 = \frac{1 + \Delta^{**}}{r} - \frac{\lambda \mu_1}{\lambda \mu_1 + \kappa + r} \frac{\Delta^{**} - \Delta^*}{r}. \quad (32)$$



This proposition shows that, the four cutoff points collapse into two:  $\Delta^*$  and  $\Delta^{**}$ . A non-owner with a type  $\Delta^*$  is indifferent from buying asset 1 and not buying any asset. A holder of asset 1 with a type  $\Delta^*$  is indifferent between holding the asset and selling it. Similarly, a non-owner with a type  $\Delta^{**}$  is indifferent from buying asset 1 and buying asset 2; a holder of asset 2 with a type  $\Delta^{**}$  is indifferent between holding the asset and selling it.

Equations (24)–(29) characterize the population size of each group. In particular, equation (24) shows that the buyers and sellers for asset 1 have the same population size. Moreover, since there is no delay in trading asset 2, at each point in time, the groups of investors who need to buy or sell asset 2 (i.e.,  $\mu_2^b$  and  $\mu_2^s$ ) are infinitesimal, as shown in equations (28) and (29). Hence, virtually all the supply of asset 2 is held by inactive holders, as shown in equation (27).

Equation (31) shows that asset 1's price has three components. The first term,  $\frac{1+\Delta^*}{r}$ , is the marginal investor's present value of the cash flow and convenience yield  $\Delta^*$  from the asset. The second term reflects the liquidity effect from the buyers, whose types range from  $\Delta^*$  to  $\Delta^{**}$ . Eager to get the asset, they are willing to pay a higher price. On the other hand, the trading friction makes sellers, whose types range from 0 to  $\Delta^*$ , willing to sell at a low price. This effect is captured by the third term. When the search friction disappears, i.e.,  $\lambda$  goes to infinity, the last two terms converge to 0 and  $P_1$  converges to  $\frac{1+\Delta^*}{r}$ .

The price of asset 2 is in equation (32). The first term,  $\frac{1+\Delta^{**}}{r}$ , is the marginal investor's present value of the cash flow and convenience yield  $\Delta^{**}$  from the asset. The second term reflects the discount due to the investors' outside option of buying asset 1. Asset 1 is cheaper, but one has to face a delay in the transaction. The higher the search friction, the less valuable the outside option of buying asset 1 is. When the search friction goes to infinity (i.e.,  $\lambda$  goes to 0), the outside option value goes to 0 and the second term becomes 0. On the other hand, when the search friction disappears, i.e.,  $\lambda$  goes to infinity,  $P_2$  converges to  $\frac{1+\Delta^*}{r}$ . That is, when the search friction disappears, the two assets become the same and have the same price.

**Proposition 2** *The effect of the search friction on asset prices is as follows:*

$$\begin{aligned} \frac{\partial P_1}{\partial \lambda} &< 0 \text{ if } \Delta^{**} - \Delta^* > \int_0^{\Delta^{**}} F(\Delta) d\Delta, \\ \frac{\partial P_1}{\partial \lambda} &> 0 \text{ if } \Delta^{**} - \Delta^* < \int_0^{\Delta^{**}} F(\Delta) d\Delta, \\ \frac{\partial P_2}{\partial \lambda} &< 0. \end{aligned}$$

When the search technology in market 1 improves, its effect on  $P_1$  depends on the tradeoff between the effect on buyers and the effect on sellers, which are captured by the second and third terms in equation (31). Note that the condition  $\Delta^{**} - \Delta^* > \int_0^{\Delta^{**}} F(\Delta) d\Delta$  is equivalent to the second term being larger than the third term, that is, the effect on buyers dominates. In this case, due to the search friction, buyers push up  $P_1$ . Hence, when the search technology improves, this effect weakens and  $P_1$  decreases. Similarly, in the other case,  $\Delta^{**} - \Delta^* < \int_0^{\Delta^{**}} F(\Delta) d\Delta$ , the effect on sellers dominates and  $P_1$  increases when the search technology improves.

Finally, when the search technology improves, it increases asset 2 buyers' outside option value, since they can more easily obtain asset 1. This reduces the comparative advantage of asset 2 and so reduces  $P_2$ .

## 2.6 The liquidity premium

Since assets 1 and 2 have identical cash flows, the price difference,  $P_2 - P_1$ , reflects the liquidity premium. From (31) and (32), the liquidity premium is given by

$$LP = \frac{\Delta^{**} - \Delta^* + \frac{\kappa}{r} \int_0^{\Delta^{**}} F(\Delta) d\Delta}{\lambda\mu_1 + \kappa + r}. \quad (33)$$

The above expression immediately shows that the liquidity premium is always positive and decreases when the search friction decreases (i.e., when  $\lambda$  increases). As  $\lambda$  goes to infinity, the friction in market 1 disappears, and the liquidity premium converges to 0.

Another observation from (33) is that the liquidity premium depends on not only the marginal investor's liquidity preference  $\Delta^{**}$ , but also the distribution of all investors' preferences  $F(\cdot)$ . In particular, the liquidity premium is increasing in  $\Delta^{**}$  but decreasing in  $\Delta^*$ . Intuitively, investor  $\Delta^{**}$  is the marginal investor who is indifferent between buying assets 1 and 2. He can pay  $P_2$

to obtain asset 2 right away. Asset 1 is cheaper, but he has to face a delay in the transaction. In the meantime, he is giving up his convenience  $\Delta^{**}$ . The investor is indifferent about the two assets if the price difference (i.e., the liquidity premium) is the same as the present value of the convenience that the marginal investor expects to lose during his search. Hence, the liquidity premium increases in  $\Delta^{**}$ .

It is less obvious that the liquidity premium also depends on  $\Delta^*$ . The intuition is the following. Suppose  $\Delta^*$  decreases. This reduces  $P_1$  since the type- $\Delta^*$  investor is the marginal investor between buying asset 1 and not buying any asset. How does  $P_2$  respond to the drop in  $P_1$ ? For investor- $\Delta^{**}$  to be indifferent between assets 1 and 2,  $P_2$  has to decrease. If  $P_1$  drops by one dollar, how much should  $P_2$  decrease to keep investor- $\Delta^{**}$  indifferent? The answer is *less* than one dollar. The reason is that, as noted in Lemma 1, an investor's expected utility is more sensitive to  $P_2$  than to  $P_1$ . That is, after a one-dollar drop in  $P_1$ , it takes a smaller drop in  $P_2$  to keep the investor indifferent between the two assets. Therefore, a decrease in  $\Delta^*$  increases the liquidity premium. The above result naturally leads to the following proposition.

**Proposition 3** *The liquidity premium decreases in  $X_2$  (i.e.,  $\frac{\partial LP}{\partial X_2} < 0$ ) if*

$$\frac{1}{f(\Delta^*)} + \frac{\lambda\kappa X_1 \left[ \Delta^{**} - \Delta^* + \frac{\kappa}{r} \int_0^{\Delta^{**}} F(\Delta) d\Delta \right]}{(2\lambda\mu_1 + \kappa)(\lambda\mu_1 + \kappa + r)} < \frac{N \left( 1 + \frac{\kappa}{r} F(\Delta^{**}) \right)}{N - X_1} \frac{1}{f(\Delta^{**})}, \quad (34)$$

*but increases in  $X_2$  (i.e.,  $\frac{\partial LP}{\partial X_2} > 0$ ) if*

$$\frac{1}{f(\Delta^*)} + \frac{\lambda\kappa X_1 \left[ \Delta^{**} - \Delta^* + \frac{\kappa}{r} \int_0^{\Delta^{**}} F(\Delta) d\Delta \right]}{(2\lambda\mu_1 + \kappa)(\lambda\mu_1 + \kappa + r)} > \frac{N \left( 1 + \frac{\kappa}{r} F(\Delta^{**}) \right)}{N - X_1} \frac{1}{f(\Delta^{**})}. \quad (35)$$

This proposition shows that the supply of asset 2 may increase or decrease the liquidity premium, depending on the distribution of the investors' liquidity preferences. Intuitively, since an increase in  $X_2$  attracts more investors with high  $\Delta$ , it pushes down both  $\Delta^*$  and  $\Delta^{**}$ . That is, the increase in  $X_2$  has two effects. First, it decreases  $\Delta^{**}$  and so decreases the premium. Second, it decreases  $\Delta^*$  and so increases the liquidity premium. The strength of the two effects depends

on the sensitivity of  $\Delta^*$  and  $\Delta^{**}$  to  $X_2$ . From (22) and (23), we have

$$\begin{aligned}\frac{\partial \Delta^*}{\partial X_2} &= -\frac{1}{Nf(\Delta^*)}, \\ \frac{\partial \Delta^{**}}{\partial X_2} &= -\frac{1}{(N - X_1)f(\Delta^{**})}.\end{aligned}$$

So, the strength of the two effects is decreasing in  $f(\Delta^*)$  and  $f(\Delta^{**})$ , respectively.

Intuitively, a higher  $f(\Delta^{**})$  means that there are more investors whose types are around  $\Delta^{**}$ . Hence, an increase in  $X_2$  pushes down  $\Delta^{**}$  less, and so the first effect (i.e., the effect through  $\Delta^{**}$ ) is weaker. Similarly, the strength of the second effect is weaker if  $f(\Delta^*)$  is larger. This is illustrated in Figure 1. Panel A reflects condition (34):  $f(\Delta^*)$  is high relative to  $f(\Delta^{**})$ . Hence, the first effect (i.e., the effect through  $\Delta^{**}$ ) dominates and the supply of asset 2 decreases the liquidity premium. Similarly, under condition (35), as illustrated in Panel B,  $f(\Delta^{**})$  is high relative to  $f(\Delta^*)$ . The second effect (i.e., the effect through  $\Delta^*$ ) dominates and an increase in  $X_2$  increases the liquidity premium.

To better illustrate the result in Proposition 3, and also demonstrate that conditions (34) and (35) are both attainable, we parameterize the density function  $f(\cdot)$  as

$$f(\Delta) = a\Delta^{a-1}, \tag{36}$$

for  $\Delta \in (0, 1)$ , where  $a$  is a constant and  $a > 0$ . The case  $a = 1$  corresponds to the uniform distribution. When  $a$  increases, the slope of  $f(\cdot)$  increases. So, a small  $a$  corresponds to the case in Panel A of Figure 1, and a large  $a$  represents the case in Panel B.

**Corollary 1** *For the distribution in (36),  $\frac{\partial LP}{\partial X_2} < 0$  if  $a < \hat{a}$ , and  $\frac{\partial LP}{\partial X_2} > 0$  if  $a > \hat{a}$ , where  $\hat{a}$  is a constant and given by equation (76) in the Appendix.*

In the uniform distribution case, i.e.,  $a = 1$ , the liquidity premium is decreasing in  $X_2$ , since we can see from the Appendix that the constant  $\hat{a}$  is larger than 2. The corollary shows that the liquidity premium becomes increasing in  $X_2$  only when the slope of  $f(\cdot)$  is sufficiently large, i.e.,  $a > \hat{a}$ , as illustrated in Panel B of Figure 1.

The empirical evidence in Krishnamurthy and Vissing-Jorgensen (2012) suggests that the

supply of Treasury securities decreases their premium. This is consistent with the implication from the case  $a < \hat{a}$  or Panel A in Figure 1. That is, the liquidity preference among investors is such that many investors have a modest convenience (i.e.,  $\Delta$ ), while some other investors have large  $\Delta$ . One can think of these investors with large  $\Delta$  as banks, which can use Treasury securities as collateral to issue checking accounts, or hedge funds that use Treasury securities as collateral for their derivative positions. Normal investors, however, do not benefit as much from the liquidity and safety in Treasury securities.

The case where  $a > \hat{a}$  (i.e., Panel B in Figure 1) may be relevant for some other occasions. For example, if one interprets asset 1 as junk bonds and asset 2 as bonds with investment grade and above, such as investment-grade corporate bonds, agency bonds and Treasury securities etc. Hence, most investors hold asset 2 for its liquidity and safety, and only a small of investors with expertise (e.g., hedge funds) are marginal investors for junk bonds. That is,  $f(\Delta^*)$  is small relative to  $f(\Delta^{**})$ , as in Panel B. In this case, the novel prediction from our model is that when the supply of Treasury or investment-grade bonds increases, the spread between junk bonds and investment-grade bonds should go up.<sup>4</sup>

## 2.7 Trading needs and asset prices

How do investors' trading needs affect the asset prices and liquidity premium? In the model, investors' trading needs are summarized by  $\kappa$ . The higher  $\kappa$  is, the more frequently each investor's type changes, and hence the stronger the trading need. From Proposition 1, we obtain the following.

### Proposition 4

$$\frac{\partial P_1}{\partial \kappa} \begin{cases} > 0 & \text{if } \Delta^{**} - \Delta^* < \int_0^{\Delta^{**}} F(\Delta) d\Delta \\ < 0 & \text{if } \Delta^{**} - \Delta^* > \int_0^{\Delta^{**}} F(\Delta) d\Delta \end{cases}$$

$$\frac{\partial P_2}{\partial \kappa} \begin{cases} < 0 & \text{if } \kappa < \kappa^*, \\ > 0 & \text{if } \kappa > \kappa^*, \end{cases}$$

---

<sup>4</sup>We run regressions similar to those in Krishnamurthy and Vissing-Jorgensen (2012). However, the high yield index is available only after 1997. Perhaps due to the short sample period, we do not find a significant relation between the Treasury supply and the spread between junk bonds and investment-grade bonds.

where

$$\kappa^* \equiv \frac{r}{1 + \sqrt{\frac{rN}{\lambda X_1(N - X_1 - X_2)}}}.$$

This proposition shows that the impact of trading need on  $P_1$  depends on the impacts of the buyers and sellers in market 1. As noted in Proposition 2,  $\Delta^{**} - \Delta^* < \int_0^{\Delta^{**}} F(\Delta) d\Delta$  implies that the buyers' impact dominates. In this case, more trading need increases  $P_1$ . Similarly, if the sellers' impact dominates, i.e.,  $\Delta^{**} - \Delta^* > \int_0^{\Delta^{**}} F(\Delta) d\Delta$ , more trading need decreases  $P_1$ .

The effect of  $\kappa$  on  $P_2$  is more subtle. When  $\kappa$  increases, it has two effects. First, it means more investors search in market 1, making it more liquid. This reduces asset 2's advantage and decreases  $P_2$ . Second, a higher  $\kappa$  also means that investors expect a shorter holding period. This makes the delay in trading asset 1 even less appealing, and hence increases  $P_2$ . When  $\kappa$  is smaller than  $\kappa^*$ , the first effect dominates and  $\frac{\partial P_2}{\partial \kappa} < 0$ . In fact, when  $\kappa$  goes to 0, both  $\mu_1^s$  and  $\mu_1^b$  go to 0, that is, market 1 becomes completely illiquid and  $\frac{\partial P_2}{\partial \kappa}$  converges to  $-\infty$ . On the other hand, when  $\kappa > \kappa^*$ , investors expect to hold an asset only for a short period of time. This makes the delay in market 1 less tolerable. Hence, the second effect dominates and  $\frac{\partial P_2}{\partial \kappa} > 0$ . Taken together, it is easy to see that the effect of  $\kappa$  on the liquidity premium is mixed and depends on the relative strength of the four effects discussed above.

## 2.8 Welfare

This section endogenizes the investment in the search technology, and analyzes the welfare implications. In particular, we specify the cost of investing in the search technology and the corresponding matching function as the following. Investor  $i$  has to pay  $\Gamma(\lambda_i)$  to obtain a search technology  $\lambda_i$ , where  $\Gamma(\cdot)$  is continuous, differentiable, increasing, and convex, with  $\Gamma(0) = 0$ ,  $\Gamma'(\infty) = \infty$ . For simplicity, the cost  $\Gamma(\lambda_i)$  is paid at  $t = 0$  before the investor knows his type, and there is no further cost to maintain the technology and investors cannot make adjustments to their technology after  $t = 0$ . Suppose investor  $i$  is a buyer in market 1. Let  $\bar{\lambda}$  denote the average technology chosen by sellers. Then, during  $[t, t + dt)$  this buyer meets a seller with a probability  $[\alpha\lambda_i + (1 - \alpha)\bar{\lambda}] \mu_1^s dt$ . That is, the matching intensity is a linear combination of the buyer's technology  $\lambda_i$  and the average technology of all sellers  $\bar{\lambda}$ . Similarly, suppose that investor

$i$  is a seller in market 1 and that  $\bar{\lambda}$  is buyers' average technology. Then, during  $[t, t + dt)$  this seller meets a buyer with a probability  $[\alpha\lambda_i + (1 - \alpha)\bar{\lambda}] \mu_1^b dt$ .

An investor's objective function is

$$\max_{\lambda_i} \mathbf{E}[V(\Delta)] - \Gamma(\lambda_i) \tag{37}$$

where  $\mathbf{E}[V(\Delta)]$  is an investor's expected value function across states in the steady states. We consider a symmetric equilibrium, in which all investors choose the same level of technology. One degenerate equilibrium is that all investors choose not to invest in their search technology at all and the market for asset 1 is shut down. In the following, we focus on the more interesting equilibrium where investors choose to invest, and denote this decentralized choice as  $\lambda^d$ .

As a comparison, we also analyze the choice of a central planner, who chooses the technology investment for all investors to maximize

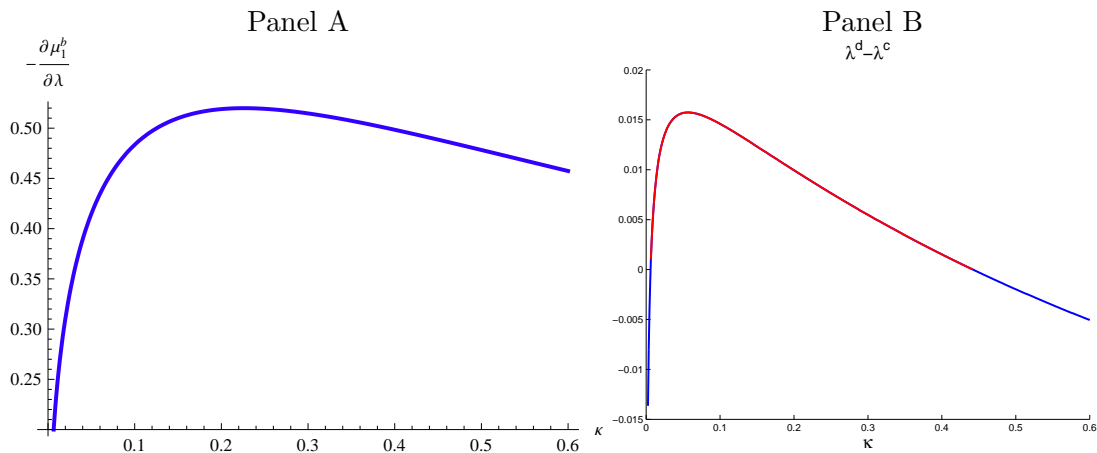
$$\max_{\lambda} \mathbf{E}[V(\Delta)] - \Gamma(\lambda). \tag{38}$$

We denote this centralized choice as  $\lambda^c$ . The difference between (37) and (38) is that when an investor makes a decentralized decision in (37), he takes other investors' choice  $\bar{\lambda}$  and the population distribution (e.g.,  $\mu_1^b$  and  $\mu_1^s$ ) as given. In (38), however, the central planner internalizes the consequences of investors' decisions. The following proposition compares the investment choices across the two cases.

**Proposition 5** *There are unique solutions  $\lambda^d$  and  $\lambda^c$  to (37) and (38), respectively. If  $\alpha \leq \frac{1}{2}$ , decentralized decisions lead to underinvest, i.e.,  $\lambda^d < \lambda^c$ . If  $\alpha > \frac{1}{2}$ , decentralized decisions may lead to over- or underinvestment.*

There are two externalities in this economy. First, an investor's investment in his technology also benefits his potential future trading partners. This positive externality leads to a free-riding problem, and hence underinvestment relative to the first best. Second, as the search technology improves, more investors' trading needs get matched, and hence fewer investors are left searching in the market, reducing the marginal benefit of searching for all investors. This negative externality leads to overinvestment.

The strength of the first externality is determined by  $\alpha$ . The smaller the  $\alpha$ , the stronger the free riding problem. The proposition shows that in the case of  $\alpha \leq \frac{1}{2}$ , the free-riding problem always dominates and leads to underinvestment relative to the central planning case. In the case of  $\alpha > \frac{1}{2}$ , however, the second externality may dominate. In particular, Panel A of Figure 3 plots the sensitivity of the population size to the search technology,  $-\partial\mu_1^b/\partial\lambda$ , against  $\kappa$ . It shows that this sensitivity is the strongest when  $\kappa$  is in the intermediate region. This is the region where the second externality is the strongest. Hence, as shown in Panel B, in the intermediate region for  $\kappa$ , we have  $\lambda^d > \lambda^c$ , i.e., investors overinvest relative to a central planner in this region. That is, decentralized decisions lead to underinvestment in the matching technology in markets where investors expect to trade very infrequently or very frequently, but lead to overinvestment in markets where the trading frequency is intermediate.



**Figure 3:** Panel A plots  $-\partial\mu_1^b/\partial\lambda$ , against  $\kappa$ . Panel B plots  $\lambda^d - \lambda^c$ , against  $\kappa$ . Parameters for both panels:  $X_1 = 10$ ,  $X_2 = 10$ ,  $N = 22$ . Other parameters for Panel B:  $\alpha = 0.7$ ,  $r = 0.02$ ,  $\bar{\Delta} = 1$ ,  $\Gamma(\lambda) = 0.1\lambda^4$ .

### 3 The safety premium

The analysis so far has focused on the liquidity premium. We now move on to analyze the safety premium. In particular, we modify the model by introducing a default risk to asset 1. Specifically, asset 1 pays a constant cash flow of \$1 per unit of time, until default, which has an intensity of  $\pi$ . That is, during  $[t, t + dt)$ , a fraction  $\pi dt$  of asset-1 holders lose their holdings in asset 1, while



the remaining asset-1 holders are intact. If default happens to an investor who is trying to sell his asset 1, he becomes an inactive non-owner. Alternatively, if an investor is an inactive holder of asset 1 when default happens to his holding, he then chooses his optimal strategy (buy asset 1, buy asset 2, or stay inactive) according to his current type  $\Delta$ .

To keep the steady state stable, we assume that  $X_1\pi dt$  units of asset 1 are issued to market 1 during  $[t, t + dt)$ , so that the total amount of asset 1 outstanding remains a constant over time. We can think of the sellers of the newly issued asset 1 as investment bankers. They are treated the same as other sellers in market 1. The only difference is that the investment bankers leave the market after they sell their assets. Hence, at each point in time, some investment bankers leave and market and other investment bankers enter the market with newly issued asset 1. In the steady state, the population size of investment bankers in the market remain constant over time. The steady-state equilibrium is defined analogously to that in Definition 1, and is characterized in the following proposition.

**Proposition 6** *The steady-state equilibrium is given by*

$$P_1 = \frac{1 + \Delta^\dagger}{\pi + r} + \frac{\kappa}{\pi + r} \frac{\Delta^{\dagger\dagger} - \Delta^\dagger - \int_0^{\Delta^\dagger} F(\Delta) d\Delta}{\lambda\mu_1^b + \kappa + \pi + r}, \quad (39)$$

$$P_2 = \frac{1 + \Delta^{\dagger\dagger}}{r} - \frac{\lambda\mu_1^b}{\lambda\mu_1^b + \kappa + \pi + r} \frac{\Delta^{\dagger\dagger} - \Delta^\dagger}{r}, \quad (40)$$

where  $\mu_1^b$  is the solution to

$$\frac{1}{\kappa} \left( \mu_1^b + \frac{\kappa + \pi}{\lambda} \right) \left[ \frac{\lambda\mu_1^b + \pi}{X_1} - \frac{\pi}{\mu_1^b} \right] = 1 - \frac{\frac{1}{\pi + \kappa} \lambda (\mu_1^b)^2 + \mu_1^b + X_2}{N - \frac{\kappa}{\pi + \kappa} \frac{\lambda\mu_1^b}{\lambda\mu_1^b + \pi} X_1}, \quad (41)$$

and

$$\begin{aligned} F(\Delta^{\dagger\dagger}) &= 1 - \frac{X_2}{N - \frac{\kappa}{\pi + \kappa} \frac{\lambda\mu_1^b}{\lambda\mu_1^b + \pi} X_1}, \\ F(\Delta^\dagger) &= \frac{1}{\kappa} \left( \mu_1^b + \frac{\kappa + \pi}{\lambda} \right) \left( \frac{\lambda\mu_1^b + \pi}{X_1} - \frac{\pi}{\mu_1^b} \right), \\ \mu_1^s &= \mu_1^b - \frac{\pi X_1}{\lambda\mu_1^b + \pi}, \\ \mu_1^h &= X_1 - \mu_1^b, \\ \mu_0^h &= N - X_2 - \frac{\lambda\mu_1^b}{\lambda\mu_1^b + \pi} X_1 - \mu_1^b. \end{aligned}$$

The equilibrium shares many similar properties to those in Proposition 1. For example, similar to the two cutoff points in the baseline model, we now have two cutoff points  $\Delta^\dagger$  and  $\Delta^{\dagger\dagger}$ . Investor- $\Delta^\dagger$  is indifferent between searching to buy asset 1 and staying inactive, and investor- $\Delta^{\dagger\dagger}$  is indifferent between searching to buy asset 1 and buying asset 2.

The price of asset 1 is determined by the valuation of the marginal investor  $\Delta^\dagger$  (i.e.,  $\frac{1+\Delta^\dagger}{\pi+r}$ ) and the illiquidity effect from the buyers and sellers (i.e., the last term in equation (39)). The price of asset 2 is determined by its marginal investor's valuation  $\frac{1+\Delta^{\dagger\dagger}}{r}$ , and the discount due to the investor's outside option of buying asset 1 (i.e., the last term in equation (40)). When the search friction disappears, i.e.,  $\lambda$  goes to infinity, asset 1 becomes perfectly liquid and its price  $P_1$  converges to  $\frac{1+\Delta^*}{\pi+r}$ , and  $P_2$  converges to  $\frac{1+\Delta^*}{r}$ .

The price difference,  $P_2 - P_1$ , is due to the better liquidity and safety of asset 2. To isolate the impact from safety, we define the safety premium as

$$SP \equiv \lim_{\pi \rightarrow 0} P_1 - P_1,$$

where  $\lim_{\pi \rightarrow 0} P_1$  is the limit of the price of asset 1 when the default intensity converges to 0. One can think of  $\lim_{\pi \rightarrow 0} P_1$  as the price of an asset that is as liquid as asset 1, but as safe as asset 2. Hence,  $SP$  reflects the safety premium that asset 2 commands. The following proposition characterizes the properties of the safety premium.

**Proposition 7** *If  $\lambda$  is sufficiently large, the safety premium decreases with the supply of asset 2,  $\frac{\partial SP}{\partial X_2} < 0$ , and this impact is stronger when the default intensity is higher,  $\frac{\partial^2 SP}{\partial X_2 \partial \pi} < 0$ .*

Due to the default risk, the expected cash flow from asset 1 is lower. So, it is not surprising that there is a safety premium. However, the above proposition shows that the safety premium is related to the supply of asset 2. Intuitively, in the absence of default, the marginal investor of asset 1 enjoys a convenience yield of  $\Delta^\dagger$ . The default risk, however, means that he can get only a fraction of it in expectation. That is, the safety premium reflects a fraction of the convenience yield  $\Delta^\dagger$  that is expected to be wiped out by default. Hence, the safety premium increases in  $\Delta^\dagger$ . When the supply of asset 2 increases, it attracts more investors with high types, and so reduces  $\Delta^\dagger$  and the safety premium. Moreover, when the default intensity  $\pi$  is higher, the safety

premium reflects a larger fraction of the convenience yields  $\Delta^\dagger$ , and hence is more sensitive to  $\Delta^\dagger$ . Therefore, the effect of supply of asset 2 on the safety premium is stronger.

## 4 Conclusion

We have analyzed a micro-founded model of the safety and liquidity premium. Relative to the reduced-form money-in-the-utility-function approach, our model explicitly examines investors' trading needs and trading frictions. One new insight from our approach is that the marginal investor's preference for safety and liquidity is no longer enough in determining the premium. Instead, the distribution of all investors' preferences plays a direct role. The model implies that an increase in the supply of Treasury securities decreases the credit spread of investment-grade bonds, but may increase the spread between junk bonds and investment-grade bonds. Our analysis highlights the importance of explicitly modeling trading frictions. This is parallel to the point stressed in the classical search-theoretical model of Kiyotaki and Wright (1989), which emphasizes the importance of explicitly modeling the frictions that render money essential.

## Appendix

In the following, we sketch the proofs of our main results and leave the details of some derivations to our online appendix at <http://faculty.som.yale.edu/hongjunyan/>.

### Proof of Propositions 1–4

**Step I.** Non-owner's optimal strategy. Equation (19) implies that  $V_0^h(\Delta)$  is a constant for all  $\Delta$ .

We denote it by  $U \equiv V_0^h(\Delta)$ . Differentiating (16) and (13), we obtain

$$\frac{dV_2^b(\Delta)}{d\Delta} = \frac{dV_2^h(\Delta)}{d\Delta} = \frac{1}{\kappa + r}, \quad (42)$$

$$\frac{dV_1^b(\Delta)}{d\Delta} = \frac{\lambda\mu_1^s}{\lambda\mu_1^s + \kappa + r} \frac{dV_1^h(\Delta)}{d\Delta} = \frac{\lambda\mu_1^s}{\lambda\mu_1^s + \kappa + r} \frac{1}{\kappa + r}. \quad (43)$$

Hence,  $V_2^b(\Delta)$  and  $V_1^b(\Delta)$  are linear in  $\Delta$  and  $\frac{dV_2^b(\Delta)}{d\Delta} > \frac{dV_1^b(\Delta)}{d\Delta} > 0 = \frac{dV_0^h(\Delta)}{d\Delta}$ , for any  $\Delta$ . We thus conjecture that there exist two cutoff points,  $\Delta_0^*$  and  $\Delta_0^{**}$ , such that

$$\max\{V_0^h(\Delta), V_1^b(\Delta), V_2^b(\Delta)\} = \begin{cases} U, & \text{if } \Delta \in [0, \Delta_0^*), \\ V_1^b(\Delta), & \text{if } \Delta \in (\Delta_0^*, \Delta_0^{**}), \\ V_2^b(\Delta), & \text{if } \Delta \in (\Delta_0^{**}, \bar{\Delta}], \end{cases} \quad (44)$$

$$V_1^b(\Delta_0^*) = V_0^h(\Delta_0^*) = U, \quad (45)$$

$$V_1^b(\Delta_0^{**}) = V_2^b(\Delta_0^{**}). \quad (46)$$

From (45) and (46), we can write  $V_1^b(\Delta)$  and  $V_2^b(\Delta)$  as

$$V_1^b(\Delta) = V_1^b(\Delta_0^*) + \frac{\lambda\mu_1^s}{\lambda\mu_1^s + \kappa + r} \frac{\Delta - \Delta_0^*}{\kappa + r} = U + \frac{\lambda\mu_1^s}{\lambda\mu_1^s + \kappa + r} \frac{\Delta - \Delta_0^*}{\kappa + r}, \quad (47)$$

$$V_2^b(\Delta) = V_2^b(\Delta_0^{**}) + \frac{\Delta - \Delta_0^{**}}{\kappa + r} = U + \frac{\lambda\mu_1^s}{\lambda\mu_1^s + \kappa + r} \frac{\Delta_0^{**} - \Delta_0^*}{\kappa + r} + \frac{\Delta - \Delta_0^{**}}{\kappa + r}, \quad (48)$$

where we have used  $V_1^b(\Delta_0^*) = U$  in (47) and  $V_2^b(\Delta_0^{**}) = U + \frac{\lambda\mu_1^s}{\lambda\mu_1^s + \kappa + r} \frac{\Delta_0^{**} - \Delta_0^*}{\kappa + r}$  in (48).

From (19) and the optimal strategy specified in (44), we have

$$U = \frac{\kappa}{\kappa + r} \left[ \int_0^{\Delta_0^*} U dF(\Delta) + \int_{\Delta_0^*}^{\Delta_0^{**}} V_1^b(\Delta) dF(\Delta) + \int_{\Delta_0^{**}}^{\bar{\Delta}} V_2^b(\Delta) dF(\Delta) \right].$$

Substituting (47) and (48) into the above equation and rearranging, we obtain

$$U = \frac{\kappa}{r} \left[ \frac{\lambda\mu_1^s}{\lambda\mu_1^s + \kappa + r} \frac{\int_{\Delta_0^*}^{\Delta_0^{**}} [1 - F(\Delta)] d\Delta}{\kappa + r} + \frac{\int_{\Delta_0^{**}}^{\bar{\Delta}} [1 - F(\Delta)] d\Delta}{\kappa + r} \right]. \quad (49)$$

**Step II.** Asset 2 owner's optimal strategy. Differentiating (18), we obtain

$$\frac{dV_2^h(\Delta)}{d\Delta} = \frac{1}{\kappa + r}. \quad (50)$$

Equations (17) and (44) imply that

$$V_2^s(\Delta) = \begin{cases} U + P_2, & \text{if } \Delta < \Delta_0^*, \\ U + P_2 + \frac{\lambda\mu_1^s}{\lambda\mu_1^s + \kappa + r} \frac{\Delta - \Delta_0^*}{\kappa + r}, & \text{if } \Delta \geq \Delta_0^*. \end{cases} \quad (51)$$

Since the slope of  $V_2^h(\Delta)$  is larger than that of  $V_2^s(\Delta)$  for all  $\Delta$ , we conjecture that there exists a cutoff point  $\Delta_2^*$  such that

$$\max\{V_2^s(\Delta), V_2^h(\Delta)\} = \begin{cases} V_2^s(\Delta), & \text{if } \Delta < \Delta_2^*, \\ V_2^h(\Delta), & \text{if } \Delta \geq \Delta_2^*, \end{cases} \quad (52)$$

$$V_2^s(\Delta_2^*) = V_2^h(\Delta_2^*). \quad (53)$$

We have the following chain of equalities:

$$\begin{aligned} U + \frac{\lambda\mu_1^s}{\lambda\mu_1^s + \kappa + r} \frac{\Delta_0^{**} - \Delta_0^*}{\kappa + r} &\stackrel{(a)}{=} V_1^b(\Delta_0^{**}) \stackrel{(b)}{=} V_2^b(\Delta_0^{**}) \stackrel{(c)}{=} V_2^h(\Delta_0^{**}) - P_2 \\ &\stackrel{(d)}{=} \frac{1 + \Delta_0^{**} + \kappa\mathbf{E}[\max\{V_2^s(\Delta'), V_2^h(\Delta')\}]}{\kappa + r} - P_2, \end{aligned}$$

where (a) is due to (47), (b) is due to (46), (c) is due to (16), and (d) is due to (18). Rearranging, we have

$$U + P_2 = \frac{1 + \Delta_0^{**} + \kappa\mathbf{E}[\max\{V_2^s(\Delta'), V_2^h(\Delta')\}]}{\kappa + r} - \frac{\lambda\mu_1^s}{\lambda\mu_1^s + \kappa + r} \frac{\Delta_0^{**} - \Delta_0^*}{\kappa + r}. \quad (54)$$

It is easy to verify that  $\Delta_2^* > \Delta_0^*$ . Hence, equation (51) implies

$$V_2^s(\Delta_2^*) = U + P_2 + \frac{\lambda\mu_1^s}{\lambda\mu_1^s + \kappa + r} \frac{\Delta_2^* - \Delta_0^*}{\kappa + r}.$$

The above equation, (18), and (53) imply

$$U + P_2 = \frac{1 + \Delta_2^* + \kappa\mathbf{E}[\max\{V_2^s(\Delta'), V_2^h(\Delta')\}]}{\kappa + r} - \frac{\lambda\mu_1^s}{\lambda\mu_1^s + \kappa + r} \frac{\Delta_2^* - \Delta_0^*}{\kappa + r}. \quad (55)$$

From equations (54) and (55) we have  $\Delta_2^* = \Delta_0^{**} \equiv \Delta^{**}$ .

**Step III.** Asset 1 owner's optimal strategy. Differentiating (14) and (15), we obtain

$$\frac{dV_1^h(\Delta)}{d\Delta} = \frac{1}{\kappa + r}, \quad (56)$$

$$\frac{dV_1^s(\Delta)}{d\Delta} = \begin{cases} \frac{1}{\lambda\mu_1^b + \kappa + r}, & \text{if } \Delta < \widehat{\Delta}_0, \\ \frac{1}{\kappa + r}, & \text{if } \Delta > \widehat{\Delta}_0, \end{cases} \quad (57)$$

where  $\widehat{\Delta}_0 = \frac{\lambda\mu_1^s\Delta_0^* + (\kappa+r)\Delta_0^{**}}{\lambda\mu_1^s + \kappa + r} \in (\Delta_0^*, \Delta_0^{**})$ .

The slope of  $V_1^s(\Delta)$  and  $V_1^h(\Delta)$  are the same for the region  $\Delta > \widehat{\Delta}_0$ . We must have  $V_1^s(\widehat{\Delta}_0) \leq V_1^h(\widehat{\Delta}_0)$ , because otherwise it implies  $V_1^s(\Delta) > V_1^h(\Delta)$  for all  $\Delta$ , i.e., no investors want to hold asset 1. Hence, there should be a cutoff point  $\Delta_1^*$ , such that  $\Delta_1^* \leq \widehat{\Delta}_0$  and

$$\max\{V_1^s(\Delta), V_1^h(\Delta)\} = \begin{cases} V_1^s(\Delta), & \text{if } \Delta < \Delta_1^*, \\ V_1^h(\Delta), & \text{if } \Delta \geq \Delta_1^*, \end{cases} \quad (58)$$

$$V_1^s(\Delta_1^*) = V_1^h(\Delta_1^*). \quad (59)$$

From (57), we obtain

$$V_1^s(\Delta) = \begin{cases} V_1^s(\Delta_1^*) + \frac{\Delta - \Delta_1^*}{\lambda\mu_1^b + \kappa + r}, & \text{if } \Delta \leq \widehat{\Delta}_0, \\ V_1^s(\Delta_1^*) + \frac{\widehat{\Delta}_0 - \Delta_1^*}{\lambda\mu_1^b + \kappa + r} + \frac{\Delta - \widehat{\Delta}_0}{\kappa + r}, & \text{if } \Delta > \widehat{\Delta}_0. \end{cases} \quad (60)$$

Since  $\Delta_1^* \leq \widehat{\Delta}_0$ , we have the following chain of equalities:

$$\begin{aligned} V_1^s(\Delta_1^*) &\stackrel{(a)}{=} \frac{\overbrace{\stackrel{(b)}{=}(\kappa+r)V_1^h(\Delta_1^*) \stackrel{(c)}{=}(\kappa+r)V_1^s(\Delta_1^*)}}{1 + \Delta_1^* + \kappa\mathbf{E}\left[\max\left\{V_1^s(\Delta'), V_1^h(\Delta')\right\}\right]} + \lambda\mu_1^b(U + P_1) \\ &= \frac{(\kappa + r)V_1^s(\Delta_1^*) + \lambda\mu_1^b(U + P_1)}{\lambda\mu_1^b + \kappa + r} \stackrel{(d)}{=} U + P_1, \end{aligned} \quad (61)$$

where (a) is due to (15), (b) is due to (14), (c) is due to (59), and (d) is the result after some algebra. Therefore, (61) and (59) lead to

$$V_1^h(\Delta_1^*) = U + P_1. \quad (62)$$

Because  $V_1^h(\Delta)$  is linear in  $\Delta$  as shown in (56), we have

$$V_1^h(\Delta_0^*) = V_1^h(\Delta_1^*) + \frac{\Delta_0^* - \Delta_1^*}{\kappa + r} = U + P_1 + \frac{\Delta_0^* - \Delta_1^*}{\kappa + r}. \quad (63)$$

On the other hand,

$$\begin{aligned} U &\stackrel{(a)}{=} V_1^b(\Delta_0^*) \stackrel{(b)}{=} \frac{\lambda\mu_1^s[V_1^h(\Delta_0^*) - P_1] + \kappa\mathbf{E}\left[\max\left\{V_1^b(\Delta'), V_2^b(\Delta'), V_0^h(\Delta')\right\}\right]}{\lambda\mu_1^s + \kappa + r} \\ &= \frac{\lambda\mu_1^s[V_1^h(\Delta_0^*) - P_1] + (\kappa + r)U}{\lambda\mu_1^s + \kappa + r} \stackrel{(d)}{=} V_1^h(\Delta_0^*) - P_1, \end{aligned} \quad (64)$$

where (a) is due to (45), (b) is due to (13), (c) is due to (15), and (d) can be obtained after some algebra. Substituting (63) into the above equation and rearranging, we find

$$\Delta_0^* = \Delta_1^* \equiv \Delta^*. \quad (65)$$

**Step IV.** From (58), we have

$$\mathbf{E} \left[ \max \left\{ V_1^s(\Delta), V_1^h(\Delta) \right\} \right] = \int_0^{\Delta^*} V_1^s(\Delta) dF(\Delta) + \int_{\Delta^*}^{\bar{\Delta}} V_1^h(\Delta) dF(\Delta). \quad (66)$$

Here,  $V_1^h(\Delta)$  can be expressed as

$$V_1^h(\Delta) = V_1^h(\Delta^*) + \frac{\Delta - \Delta^*}{\kappa + r}.$$

Substituting the above expression and (60) into (66), after some algebra, we have

$$\mathbf{E} \left[ \max \left\{ V_1^s(\Delta), V_1^h(\Delta) \right\} \right] = V_1^h(\Delta^*) - \frac{\int_0^{\Delta^*} F(\Delta) d\Delta}{\lambda\mu_1^b + \kappa + r} + \frac{\int_{\Delta^*}^{\bar{\Delta}} [1 - F(\Delta)] d\Delta}{\kappa + r}. \quad (67)$$

From (14), we have

$$V_1^h(\Delta^*) = \frac{1 + \Delta^* + \kappa \mathbf{E} \left[ \max \left\{ V_1^s(\Delta'), V_1^h(\Delta') \right\} \right]}{\kappa + r}, \quad (68)$$

Substituting this into (67) and rearranging, we have

$$\mathbf{E} \left[ \max \left\{ V_1^s(\Delta), V_1^h(\Delta) \right\} \right] = \frac{\kappa + r}{r} \left[ \frac{1 + \Delta^*}{\kappa + r} - \frac{\int_0^{\Delta^*} F(\Delta) d\Delta}{\lambda\mu_1^b + \kappa + r} + \frac{\int_{\Delta^*}^{\bar{\Delta}} [1 - F(\Delta)] d\Delta}{\kappa + r} \right].$$

The above equation, (49), (62), and (68) imply

$$P_1 = \frac{1 + \Delta^*}{r} + \frac{\kappa}{r} \left[ \frac{\int_{\Delta^*}^{\Delta^{**}} [1 - F(\Delta)] d\Delta}{\lambda\mu_1 + \kappa + r} - \frac{\int_0^{\Delta^*} F(\Delta) d\Delta}{\lambda\mu_1 + \kappa + r} \right].$$

**Step V.** From (52), we have

$$\mathbf{E} \left[ \max \left\{ V_2^s(\Delta), V_2^h(\Delta) \right\} \right] = \int_0^{\Delta^{**}} V_2^s(\Delta) dF(\Delta) + \int_{\Delta^{**}}^{\bar{\Delta}} V_2^h(\Delta) dF(\Delta).$$

From (53) and (51), we obtain

$$V_2^h(\Delta) = U + P_2 + \frac{\lambda\mu_1^s}{\lambda\mu_1^s + \kappa + r} \frac{\Delta^{**} - \Delta^*}{\kappa + r} + \frac{\Delta - \Delta^{**}}{\kappa + r}.$$

Substituting the above equation and  $V_2^s(\Delta)$  in (51), we obtain

$$\mathbf{E}\left[\max\left\{V_2^s(\Delta), V_2^h(\Delta)\right\}\right] = U + P_2 + \frac{\lambda\mu_1^s}{\lambda\mu_1^s + \kappa + r} \frac{\int_{\Delta^*}^{\Delta^{**}} [1 - F(\Delta)] d\Delta}{\kappa + r} + \frac{\int_{\Delta^{**}}^{\bar{\Delta}} [1 - F(\Delta)] d\Delta}{\kappa + r}.$$

The above equation, (54), and (49) imply

$$P_2 = \frac{1 + \Delta^{**}}{r} - \frac{\lambda\mu_1}{\lambda\mu_1 + \kappa + r} \frac{\Delta^{**} - \Delta^*}{r}. \quad (69)$$

**Step VI.** Substituting (21) into (5) and (7), we have

$$\begin{aligned} \kappa\left(\mu_0^h + X_2 + \mu_1^b\right) F(\Delta^{**}) - \kappa\mu_1^b &= \lambda\mu_1^b\mu_1^s + \kappa\left(\mu_0^h + \mu_1^b + X_2\right) F(\Delta^*), \\ \kappa\mu_0^h &= \lambda\mu_1^b\mu_1^s + \kappa\left(\mu_0^h + \mu_1^b + X_2\right) F(\Delta^*). \end{aligned} \quad (70)$$

The above two equations imply

$$F(\Delta^{**}) = 1 - \frac{X_2}{\mu_0^h + X_2 + \mu_1^b}.$$

Substituting (12) and (20) into the above equation, we have (23). Substituting (12) and (20) in (70), we obtain

$$\kappa(N - X_1 - X_2) - \kappa\mu_1 = \lambda(\mu_1)^2 + \kappa(N - X_1) F(\Delta^*). \quad (71)$$

From (20) and (6), we have

$$\kappa\mu_1 + \lambda(\mu_1)^2 = \kappa X_1 F(\Delta^*). \quad (72)$$

From (71) and (72), we obtain a quadratic equation of  $\mu_1$ :

$$(\mu_1)^2 + \frac{\kappa}{\lambda}\mu_1 - \frac{\kappa X_1}{\lambda} \left(1 - \frac{X_1 + X_2}{N}\right) = 0. \quad (73)$$

The positive root is (30). Substituting it into (71), we have (22). It is straightforward to obtain the comparative statics in Propositions 2–4.

## Proof of Corollary 1

With  $f(\cdot)$  in (36), Proposition 3 implies that  $LP$  is increasing in  $X_2$  if and only if

$$\frac{\frac{1}{a} - B}{N^{\frac{1}{a}}} > \frac{\frac{1}{a} - B + \frac{(1-B)a+1}{a(a+1)} \frac{\kappa}{r} F(\Delta^{**})}{(N - X_1)^{\frac{1}{a}}}, \quad (74)$$



where  $B \in (0, \frac{1}{2})$  is given by

$$B = \frac{\frac{\lambda\kappa X_1}{2} F(\Delta^*)}{\left(\frac{\kappa}{2}\right)^2 + \lambda\kappa X_1 F(\Delta^*) + \left(\frac{\kappa}{2} + r\right) \sqrt{\left(\frac{\kappa}{2}\right)^2 + \lambda\kappa X_1 F(\Delta^*)}}.$$

There are 3 cases. Case 1: If  $a < \frac{1}{B}$ , (74) can be rewritten as

$$\frac{\frac{1}{a} - B}{\frac{1}{a} - B + \frac{(1-B)a+1}{a(a+1)} \frac{\kappa}{r} F(\Delta^{**})} > \frac{N^{\frac{1}{a}}}{(N - X_1)^{\frac{1}{a}}}.$$

The left hand side (LHS) of the above inequality is smaller than 1, while the right hand side (RHS) is larger than 1. So, the inequality never holds and  $LP$  is decreasing in  $X_2$ .

Case 2: If  $\frac{1}{B} \leq a < a_1$ , where  $a_1$  is given by

$$a_1 = \frac{1-B}{2B} \left(1 + \frac{\kappa}{r} F(\Delta^{**})\right) + \sqrt{\left(\frac{1-B}{2B}\right)^2 \left(1 + \frac{\kappa}{r} F(\Delta^{**})\right)^2 + \frac{1}{B} \left(1 + \frac{\kappa}{r} F(\Delta^{**})\right)},$$

the LHS of (74) is negative while the RHS of (74) is positive, so the inequality never holds. Therefore,  $LP$  is decreasing in  $X_2$ .

Case 3: If  $a \geq a_1$ , (74) holds if and only if

$$\frac{N - X_1}{N} < \left[1 - \frac{(1-B)a+1}{(a+1)(aB-1)} \frac{\kappa}{r} F(\Delta^{**})\right]^a, \quad (75)$$

Note that the LHS of (75) is between 0 and 1. The RHS of (75) is increasing in  $a$ . Moreover,  $RHS = 0$  if  $a = a_1$  and  $RHS > 1$  if  $a$  is sufficiently large. Hence, there exists a unique  $\hat{a} > a_1$  such that

$$\frac{N - X_1}{N} = \left[1 - \frac{(1-B)\hat{a}+1}{(\hat{a}+1)(\hat{a}B-1)} \frac{\kappa}{r} F(\Delta^{**})\right]^{\hat{a}} \quad (76)$$

and inequality (75) holds if and only if  $a > \hat{a}$ .

Therefore, combining all three cases, we obtain that the liquidity premium is decreasing in  $X_2$  for  $a < \hat{a}$  and increasing in  $X_2$  for  $a > \hat{a}$ .

## Proof of Proposition 5

We first compute an investor's average value function across  $\Delta$  in the steady state. For this, we use  $g_i^x(\Delta)$ , where  $x = b, s, h$  and  $i = 0, 1, 2$ , to denote the density of investors with value function  $V_i^x(\Delta)$ . Since one can sell or buy asset 2 immediately, we have  $g_2^b(\Delta) = o(1)$  and

$g_2^s(\Delta) = o(1)$  for all  $\Delta$ . In the steady state, the density and value function for other types of investors are the following: i) inactive non-owners:  $V_0^h(\Delta) = U$  is given by (49) and  $g_0^h(\Delta) = (N - X_1 - X_2 - \mu_1) \frac{f(\Delta)}{F(\Delta^*)}$  for  $\Delta \in [0, \Delta^*]$ ; ii) buyers of asset 1:  $V_1^b(\Delta)$  is given by (47) and  $g_1^b(\Delta) = \mu_1 \frac{f(\Delta)}{F(\Delta^{**}) - F(\Delta^*)}$  for  $\Delta \in [\Delta^*, \Delta^{**}]$ ; iii) inactive owners of asset 1:  $V_1^h(\Delta) = U + P_1 + \frac{\Delta - \Delta^*}{\kappa + r}$  for  $\Delta \in [\Delta^*, \Delta^{**}]$ , and

$$g_1^h(\Delta) = \begin{cases} \left[ N - \frac{\mu_1}{F(\Delta^{**}) - F(\Delta^*)} \right] f(\Delta), & \text{for } \Delta \in [\Delta^*, \Delta^{**}], \\ X_1 f(\Delta), & \text{for } \Delta \in [\Delta^{**}, \bar{\Delta}]; \end{cases}$$

iv) sellers of asset 1:  $V_1^s(\Delta) = U + P_1 + \frac{\Delta - \Delta^*}{\lambda \mu_1 + \kappa + r}$  for  $\Delta \in [\underline{\Delta}, \Delta^*]$  and  $g_1^s(\Delta) = \mu_1 \frac{f(\Delta)}{F(\Delta^*)}$  for  $\Delta \in [\underline{\Delta}, \Delta^*]$ ; v) owners of asset 2:  $V_2^h(\Delta) = U + P_2 + \frac{\Delta - \Delta^*}{\kappa + r} - \frac{\Delta^{**} - \Delta^*}{\lambda \mu_1 + \kappa + r}$  and  $g_2^h(\Delta) = (N - X_1) f(\Delta)$  for  $\Delta \in [\Delta^{**}, \bar{\Delta}]$ . The expected welfare is given by

$$\begin{aligned} \mathbf{E}[V(\Delta)] &= \frac{1}{N} \left[ \int_0^{\Delta^*} V_0^h(\Delta) g_0^h(\Delta) d\Delta + \int_{\Delta^*}^{\Delta^{**}} V_1^b(\Delta) g_1^b(\Delta) d\Delta \right. \\ &\quad \left. + \int_{\underline{\Delta}}^{\Delta^*} V_1^s(\Delta) g_1^s(\Delta) d\Delta + \int_{\Delta^*}^{\bar{\Delta}} V_1^h(\Delta) g_1^h(\Delta) d\Delta + \int_{\Delta^{**}}^{\bar{\Delta}} V_2^h(\Delta) g_2^h(\Delta) d\Delta \right] \\ &= \frac{1}{r} \left[ \frac{X_1 + X_2}{N} + \int_{\Delta^*}^{\bar{\Delta}} \Delta dF(\Delta) \right] - \frac{\kappa}{r} I_1 + \mu_1 I_2, \end{aligned} \quad (77)$$

where

$$\begin{aligned} I_1 &= \left( 1 - \frac{X_1}{N} \right) \int_{\Delta^*}^{\Delta^{**}} [F(\Delta^{**}) - F(\Delta)] d\Delta + \frac{X_1}{N} \int_0^{\Delta^*} F(\Delta) d\Delta, \\ I_2 &= \frac{1}{N} \left[ \int_{\Delta^*}^{\Delta^{**}} \frac{F(\Delta) - F(\Delta^*)}{F(\Delta^{**}) - F(\Delta^*)} d\Delta + \int_{\underline{\Delta}}^{\Delta^*} \frac{F(\Delta)}{F(\Delta^*)} d\Delta \right]. \end{aligned}$$

Note that the first term in (77) is the expected utility with no friction, and the second term is the welfare loss due to search friction. Since  $\mu_1$  itself is also a function of  $\lambda$ , we will use the notation  $\mu_1(\lambda)$  to make it explicit. We define the following function

$$G(x, y) \equiv -\frac{\frac{\kappa}{r} I_1 + I_2 y}{x + \kappa + r}, \text{ for } x > 0, y > 0,$$

and so the welfare loss is  $-G(\lambda \mu_1(\lambda), \mu_1(\lambda))$ . It is easy to see that  $G(\lambda \mu_1(\lambda), \mu_1(\lambda))$  is strictly increasing in  $\lambda$  and strictly concave in  $\lambda$  and converges to zero when  $\lambda \rightarrow \infty$ .

The optimization problem (37) is equivalent to

$$\max_{\lambda_i} G([\alpha \lambda_i + (1 - \alpha) \bar{\lambda}] \mu_1(\bar{\lambda}), \mu_1(\bar{\lambda})) - \Gamma(\lambda_i).$$

The decentralized choice  $\lambda^d$  is characterized by the following first order condition:

$$\alpha\mu_1(\lambda^d) \frac{\partial G}{\partial x}(\lambda^d\mu_1(\lambda^d), \mu_1(\lambda^d)) = \Gamma'(\lambda^d). \quad (78)$$

The optimization problem (38) is equivalent to

$$\max_{\lambda_i} G(\lambda\mu_1(\lambda), \mu_1(\lambda)) - \Gamma(\lambda).$$

The centralized choice  $\lambda^c$  is characterized by the following first order condition:

$$\left[ \frac{\partial G(\lambda\mu_1(\lambda), \mu_1(\lambda))}{\partial x} \frac{d[\lambda\mu_1(\lambda)]}{d\lambda} + \frac{\partial G(\lambda\mu_1(\lambda), \mu_1(\lambda))}{\partial y} \frac{d\mu_1(\lambda)}{d\lambda} \right] \Big|_{\lambda=\lambda^c} = \Gamma'(\lambda^c). \quad (79)$$

Define

$$\begin{aligned} H(\lambda) &\equiv \alpha\mu_1(\lambda) \frac{\partial G}{\partial x}(\lambda\mu_1(\lambda), \mu_1(\lambda)) = \alpha\mu_1(\lambda) \frac{-G(\lambda\mu_1(\lambda), \mu_1(\lambda))}{\lambda\mu_1(\lambda) + \kappa + r}, \\ K(\lambda) &\equiv \frac{\mu_1(\lambda)}{2} \frac{\lambda\mu_1(\lambda) + \kappa}{\lambda\mu_1(\lambda) + \frac{\kappa}{2}} \frac{-G(\lambda\mu_1(\lambda), \mu_1(\lambda))}{\lambda\mu_1(\lambda) + \kappa + r} + \frac{(\mu_1(\lambda))^2}{2\lambda\mu_1(\lambda) + \kappa} I_2, \end{aligned}$$

then (79) and (78) can be rewritten as

$$\begin{aligned} H(\lambda^d) &= \Gamma'(\lambda^d), \\ K(\lambda^c) &= \Gamma'(\lambda^c), \end{aligned}$$

and

$$K(\lambda) - H(\lambda) = \frac{\mu_1(\lambda)}{[\lambda\mu_1(\lambda) + \kappa + r]^2} J(\lambda),$$

where

$$J(\lambda) \equiv \left[ \frac{1}{2} - \alpha + \frac{\frac{\kappa}{2}}{2\lambda\mu_1(\lambda) + \kappa} \right] \frac{\kappa}{r} I_1 + \left[ 1 - \alpha + \frac{\kappa + r}{2\lambda\mu_1(\lambda) + \kappa} \right] \mu_1(\lambda) I_2.$$

It can be shown that  $J(\lambda)$  is decreasing in  $\lambda$ , and

$$\begin{aligned} J(\lambda)|_{\lambda=0} &= (1 - \alpha) \frac{\kappa}{r} I_1 + \left( 2 - \alpha + \frac{r}{\kappa} \right) I_2 X_1 F(\Delta^*) > 0, \\ J(\lambda)|_{\lambda=\infty} &= \left( \frac{1}{2} - \alpha \right) \frac{\kappa}{r} I_1. \end{aligned}$$

If  $\alpha \leq \frac{1}{2}$ , then  $J(\lambda)|_{\lambda=\infty} > 0$  and  $J(\lambda) > J(\lambda)|_{\lambda=\infty} > 0$  for any finite  $\lambda$  because  $J(\lambda)$  is decreasing in  $\lambda$ . That is,  $K(\lambda) > H(\lambda)$  for any finite  $\lambda$ . In this case, we have  $\lambda^c > \lambda^d$ . To see

this, we suppose  $\lambda^c < \lambda^d$ . We then have

$$\Gamma'(\lambda^d) = H(\lambda^d) < H(\lambda^c) < K(\lambda^c) = \Gamma'(\lambda^c).$$

However,  $\Gamma''(\cdot) > 0$  implies that  $\Gamma'(\lambda^c) < \Gamma'(\lambda^d)$ . This results in a contradiction.

For the case of  $\alpha > \frac{1}{2}$ , Figure 3 is sufficient to show that both over- and underinvestment are possible. In the online appendix, we characterize the necessary and sufficient condition for over- and underinvestment.

### Proof of Proposition 6

The proof of this proposition is parallel to that of Proposition 1. The extra feature is that during each instant, a fraction  $\pi X_1 dt$  of asset 1 is wiped out, and the same amount of asset 1 is issued into the economy. Instead of the quadratic equation (73) in the baseline model, we now have a more general equation (41) to determine  $\mu_1^b$ . Details for the calculations are in the online appendix.

### Proof of Proposition 7

From (41), we expand  $\mu_1^b$  as

$$\mu_1^b = m_1^b / \sqrt{\lambda} + o(1/\sqrt{\lambda}), \quad (80)$$

where

$$m_1^b = \sqrt{X_1 \left[ \pi + \kappa \left( 1 - \frac{X_1 + X_2}{N} \right) \right]}.$$

From (80), we can obtain

$$\begin{aligned} \Delta^\dagger &= \Delta^* + o(1/\sqrt{\lambda}), \\ \Delta^{\dagger\dagger} &= F^{-1} \left( 1 - \frac{X_2}{N - \frac{\kappa}{\pi + \kappa} X_1} \right) + o(1), \end{aligned}$$

where  $\Delta^*$  is given by (22). We can thus expand  $P_1$  and the safety premium as

$$\begin{aligned} P_1 &= \frac{1 + \Delta^\dagger}{\pi + r} + o(1), \\ SP &= \frac{\pi(1 + \Delta^\dagger)}{r(\pi + r)} + o(1). \end{aligned}$$

Therefore, when  $\lambda$  is sufficiently large, we have

$$\begin{aligned}\frac{\partial SP}{\partial X_2} &= -\frac{\pi}{r(\pi+r)Nf(\Delta^\dagger)} < 0, \\ \frac{\partial^2 SP}{\partial X_2 \partial \pi} &= -\frac{1}{(\pi+r)^2 Nf(\Delta^\dagger)} < 0.\end{aligned}$$

## References

- Afonso, Gara and Ricardo Lagos, 2015, Trade Dynamics in the Market for Federal Funds, *Econometrica*, forthcoming.
- Bansal, Ravi, and John Coleman, 1996, A Monetary Explanation of the Equity Premium, Term Premium, and Risk-Free Rate Puzzles, *Journal of Political Economy*, 104, 1135–1171.
- Duffie, Darrell, Nicolae Garleanu, and Lasse Pedersen, 2002, Securities Lending, Shorting and Pricing, *Journal of Financial Economics*, 66, 307–339.
- Duffie, Darrell, Nicolae Garleanu, and Lasse Pedersen, 2005, Over-the-Counter Markets, *Econometrica*, 73, 1815–1847.
- Duffie, Darrell, Nicolae Garleanu, and Lasse Pedersen, 2007, Valuation in Over-the-Counter Markets, *Review of Financial Studies*, 66, 307–339.
- Duffie, Darrell, Piotr Dworczak, and Haoxiang Zhu, 2014, Benchmarks in Search Markets, working paper.
- Feldhutter, Peter, 2012, The same bond at different prices: Identifying search frictions and selling pressures, *Review of Financial Studies* 25, 1155–1206.
- Garleanu, Nicolae, 2009, Portfolio choice and pricing in illiquid markets, *Journal of Economic Theory*, 144, 532–564.
- Gavazza, Alessandro, 2011, Leasing and secondary markets: Theory and evidence from commercial aircraft, *Journal of Political Economy*, 119, 325–377.
- Gorton, Gary, 2010, Slapped by the Invisible Hand: The Panic of 2007.
- He, Zhiguo, and Konstantin Milbradt, 2013, Endogenous Liquidity and Defaultable Debt, *Econometrica*, forthcoming.
- Hugonnier, Julien, Benjamin Lester, and Pierre-Olivier Weill, 2014, Heterogeneity in Decentralized Asset Markets, working paper.
- Kiyotaki, Nobuhiro, and Randall Wright, 1989, On Money as a Medium of Exchange, *Journal of Political Economy*, 97, 927–954.
- Kiyotaki, Nobuhiro and Randall Wright, 1993, A search-theoretic approach to monetary economics, *American Economic Review*, 83, 63–77.
- Krishnamurthy, Arvind, and Annette Vissing-Jorgensen, 2012, The Aggregate Demand for Treasury Debt, *Journal of Political Economy*, 120, 233–267.
- Lagos, Ricardo, 2010, Asset Prices and Liquidity in an Exchange Economy, *Journal of Monetary Economics*, 57, 913–930.
- Lagos Ricardo, and Guillaume Rocheteau, 2009, Liquidity in Asset Markets with Search Frictions, *Econometrica*, 77, 403–426.
- Lagos, Ricardo, Guillaume Rocheteau, and Pierre-Olivier Weill, 2011, Crises and Liquidity in OTC markets, *Journal of Economic Theory*, 146, 2169–2205.

- Lagos, Ricardo and Randall Wright, 2005, A unified framework for monetary theory and policy analysis, *Journal of political Economy*, 113, 463–484.
- Lagos, Ricardo, and Shengxing Zhang, 2014, Monetary Exchange in Over-the-Counter Markets: A Theory of Speculative Bubbles, the Fed Model, and Self-fulfilling Liquidity Crises, working paper.
- Lester, Benjamin, Andrew Postlewaite, and Randall Wright, 2012, Information, Liquidity, Asset Prices, and Monetary Policy, *Review of Economic Studies*, 79, 1209–1238.
- Lester, Benjamin, Guillaume Rocheteau, and Pierre-Olivier Weill, 2014, Competing for order flow in OTC markets, working paper.
- Li, Yiting, Guillaume Rocheteau, and Pierre-Olivier Weill, 2012, Liquidity and the threat of fraudulent assets, *Journal of Political Economy*, 120, 815–846.
- Neklyudov, Artem, 2014, Bid-Ask Spreads and the Over-the-Counter Interdealer Markets: Core and Peripheral Dealers, working paper.
- Pagnotta, Emiliano and Thomas Philippon, 2013, Competing on speed, working paper.
- Sidrauski, Miguel, 1967, Rational Choice and Patterns of Growth in a Monetary Economy, *A.E.R. Papers and Proceedings*, 57, 534–44.
- Shen, Ji, Bin Wei, and Hongjun Yan, 2015, Financial Intermediation Chains in an OTC Market, working paper.
- Trejos, Alberto and Randall Wright, 2014, Search-based models of money and finance: An integrated approach, *Journal of Economic Theory*, forthcoming.
- Vayanos, Dimitri, and Tan Wang, 2007, Search and Endogenous Concentration of Liquidity in Asset Markets, *Journal of Economic Theory*, 66, 307–339.
- Vayanos, Dimitri, and Pierre-Olivier Weill, 2008, A Search-Based Theory of the On-the-run Phenomenon, *Journal of Finance*, 63, 1361–1398.
- Vayanos, Dimitri, and Jean-Luc Vila, 2009, A Preferred-Habitat Model of the Term-Structure of Interest Rates, working paper.
- Weill, Pierre-Olivier, 2007, Leaning Against the Wind, *Review of Economic Studies*, 74, 1329–1354.
- Weill, Pierre-Olivier, 2008, Liquidity Premia in Dynamic Bargaining Markets, *Journal of Economic Theory*, 140, 66–96.
- Williamson, Stephen, and Randall Wright, 2010, New Monetarist Economics: Methods, *St. Louis Federal Reserve Bank Review*, 92, 265–302.
- Williamson, Stephen, and Randall Wright, 2011, New Monetarist Economics: Models, in *Handbook of Monetary Economics*, vol. 3A, B. Friedman and M. Woodford, eds., Elsevier, 25–96.
- Zhu, Haoxiang, 2012, Finding a Good Price in Opaque Over-the-Counter Markets, *Review of Financial Studies*, 25 1255–1285.