A Search-Based Theory of the On-the-Run Phenomenon

Dimitri Vayanos and Pierre-Olivier Weill*

April 23, 2005

Abstract

We propose a model in which assets with identical cash flows can trade at different prices. Agents enter into an infinite-horizon, steady-state market to establish long or short positions. Both the spot and the asset-lending market operate through search. Short-sellers can endogenously concentrate in one asset because of search externalities and the constraint that they must deliver the asset they borrowed. As a result, that asset enjoys both greater liquidity, measured by search times, and a higher lending fee ("specialness"). Liquidity and specialness translate into price premia that are consistent with no-arbitrage. We derive closed-form solutions for small frictions, and can generate price differentials in line with observed on-the-run premia.

*Vayanos is from the London School of Economics, London WC2A 2AE, UK, email d.vayanos@lse.ac.uk, and Weill is from the Stern School of Business, New York University, New York NY 10012, USA, email pweill@stern.nyu.edu. We thank Tobias Adrian, Yakov Amihud, Hal Cole, Darrell Duffie, Mark Fisher, Joel Hasbrouck, Kenneth Garbade, Nicolae Garleanu, Jeremy Graveline, Anna Pavlova, Lasse Pedersen, Matt Richardson, Bill Silber, Stijn Van Nieuwerburgh, Robert Whitelaw, and seminar participants at LSE, New Orleans, NY Fed, NYU, and Oxford for helpful comments.

1 Introduction

In fixed-income markets some bonds trade at lower yields than others with almost identical cash flows. In the US, for example, just-issued ("on-the-run") Treasury bonds trade at lower yields than previously issued ("off-the-run") bonds maturing on nearby dates. Warga [1992] reports that an on-the-run portfolio returns on average 55bp below an off-the-run portfolio with matched duration. Similar phenomena exist in other countries. In Japan, for example, one "benchmark" government bond trades at a yield of 60bp below other bonds with comparable characteristics.[1]

How can the yields of bonds with almost identical cash flows differ by more than 50bp? Financial economists have suggested two apparently distinct hypotheses. First, on-the-run bonds are more valuable because they are significantly more liquid than their off-the-run counterparts. Second, on-the-run bonds constitute better collateral for borrowing money in the repo market. Namely, loans collateralized by on-the-run bonds offer lower interest rates than their off-the-run counterparts, a phenomenon referred to as "specialness."[2] These hypotheses, however, can provide only a partial explanation of the on-the-run phenomenon: one must still explain why assets with almost identical cash flows can differ in liquidity and specialness.

In this paper we propose a theory of the on-the-run phenomenon. We take the view that liquidity and specialness are not independent explanations of this phenomenon, but can be explained simultaneously by short-selling activity. We determine liquidity and specialness endogenously, explain why they can differ across otherwise identical assets, and study their effect on prices. A calibration of our model for plausible parameter values can generate effects of the observed magnitude.

Our theory is based on the notion that trade in government-bond markets is bilateral and can involve search. The assumption of bilateral trade captures the over-the-counter structure of these markets: transactions between dealers and their customers are negotiated over the phone, and dealers often negotiate bilaterally in the inter-dealer market.[3] The assumption of search might seem

[1] For US evidence, see also Amihud and Mendelson [1991], Krishnamurthy [2002], Goldreich, Hanke and Nath [2002], and Strebulaev [2002]. For Japan, see Mason [1987], Boudoukh and Whitelaw [1991], and Boudoukh and Whitelaw [1993].

[2] On liquidity, Sundaresan [2002] reports that trading volume of on-the-run bonds is about ten times larger than that of off-the-run bonds, and Fleming [2002] reports that bid-ask spreads of off-the-run bills are about five times larger than when these bills are on-the-run. Specialness is measured by comparing a bond's repo rate, which is the interest rate on a loan colateralized by the bond, to the general collateral rate, which is the highest quoted repo rate. Duffie [1996] reports an average specialness of 66bp for on-the-run bonds and 26bp for their off-the-run counterparts.

[3] In the US, inter-dealer trading is conducted through brokers. Some brokers operate automated trading systems, structured as electronic limit-order books. Other brokers, however, operate voice-based systems in which orders are negotiated over the phone. Barclay, Hendershott and Kotz [2004] report that automated systems account for about 85% of trading volume for on-the-run bonds, but the situation is reversed for off-the-run bonds. To explain this

questionable given that government bonds are among the most liquid assets. On-the-run bonds are indeed very liquid, with many transactions being executed almost instantly. Transactions in off-the-run bonds, however, can take significantly more time. In particular, it can be difficult to locate a large quantity of a specific off-the-run issue. Thus, while an investor needing to buy the issue can easily contact a dealer, the dealer might not have the issue in inventory and could take time to locate it. Another illustration of search frictions is the phenomenon of "fails," whereby traders do not deliver a bond they have sold or borrowed by the time they must settle. One source of fails is the difficulty to locate the bond.[4]

We consider an infinite-horizon, steady-state economy with two risky assets paying the same cash flow. Trade occurs because agents experience hedging needs to hold long or short positions. Upon experiencing a need to hold a long position, an agent enters the market seeking to buy one of the assets. He then holds the asset until the hedging need disappears, and then seeks to sell. During the time he is holding the asset, he can lend it to a short-seller for a fee. This corresponds to a repo transaction in our model, and the fee to repo specialness.[5] Conversely, upon experiencing a need to hold a short position, an agent enters the market seeking to borrow one of the assets. She then seeks to sell the asset, and when the hedging need disappears, she seeks to buy the same asset back and return it to the lender. Both the spot and the repo market operate through search and bilateral bargaining. For simplicity, we abstract away from dealers and adopt the standard search framework (e.g., Diamond [1982]) where agents search for counterparties directly.[6]

Our model has multiple equilibria: a symmetric one where short-sellers borrow both assets, and asymmetric ones where they concentrate in one asset, declining any opportunities to borrow the other. This is because of search externalities. The more agents short an asset, the greater the asset's seller pool becomes. The asset's buyer pool also increases because of the short-sellers who need to buy the asset back. A larger buyer and seller pool implies lower search times, and the enhanced liquidity attracts more short-sellers. Thus, our theory can explain differences in liquidity between otherwise identical assets, consistent with the on-the-run phenomenon.

While the general notion of search externalities is well-understood, its application to the on-the-run phenomenon is subtle. Absent the short-sellers, there would be no differences in liquidity.

phenomenon, they propose a search-based model.

[4]Fleming and Garbade [2002] report that fails in the US market averaged $7.3 billion per day during the first eight months of 2001. For comparison, the daily volume during that period was about $250 billion (Fleming [2003]).

[5]We describe repo transactions at the beginning of Section 4. See also Duffie [1996] and Fisher [2002], among others, for more detailed descriptions.

[6]Of course, the search framework is only an idealization of price formation in actual bond markets - but so is the Walrasian auction. We expand on this point and provide further arguments in support of the search framework in Section 4.1.

Indeed, assets would have a common buyer pool, consisting of the agents seeking to establish long positions. Therefore, they would be equally easy to sell. The same would hold even with short-sellers, if these were allowed to deliver any asset and not necessarily the one they borrowed. The delivery constraint effectively "locks" short-sellers into buying one asset, thus generating asset-specific buyer pools.[7]

In the asymmetric equilibria, assets differ not only in liquidity but also in specialness. Indeed, because of the search friction, asset lenders can extract some of the short-sellers' surplus. This surplus is positive only for the more liquid asset because it is the one that short-sellers are willing to borrow. Therefore, only that asset commands a positive fee and hence is "on special." The fee constitutes an additional cash flow derived from the asset, raising its price by a specialness premium. This premium is above and beyond the one associated to the asset's superior liquidity. We show that the existence of the two premia is consistent with no-arbitrage: agents cannot profit by buying one asset and shorting the other.

While our theory can identify a liquidity and a specialness component of the on-the-run premium, it also implies that this decomposition should be interpreted with caution. Indeed, since short-sellers are attracted to the more liquid asset, the asset's specialness is partly generated from liquidity. Therefore, the specialness premium can be viewed as an additional liquidity premium. In fact, our theory implies that liquidity and specialness are linked in an even more fundamental manner because they are both generated by short-selling activity.

A calibration of our model can generate price effects of the observed magnitude even for very short search times. We show that the liquidity premium is small, and the effects are mostly generated by the specialness premium. Of course, this does not mean that liquidity does not matter; it rather means that liquidity can have large effects through specialness.

Our theory sheds light on several additional aspects of the on-the-run phenomenon. One puzzling aspect is that off-the-run bonds are viewed by traders as "scarce" and hard to locate, while at the same time being cheaper than on-the-run bonds. In our model, off-the-run bonds are indeed scarce from the viewpoint of short-sellers searching to buy and deliver them. Because, however, scarcity drives short-sellers away from these bonds, it makes them less liquid and less attractive to marginal buyers who are the agents seeking to establish long positions. Our theory also shows that when assets' issue sizes are sufficiently different the equilibrium becomes unique,

[7]The delivery constraint is prevalent in actual markets, as can be seen, for example, by the incidence of short-squeezes. In a short-squeeze, short-sellers have difficulty delivering the asset they borrowed and the asset's specialness in the repo market increases dramatically. For a description of short-squeezes see, for example, Dupont and Sack [1999].

with short-sellers concentrating in the largest-supply asset. Therefore, if one views off-the-run bonds as being in smaller effective supply,[8] our theory suggests that they are less likely to attract short-sellers and thus less liquid. Finally, our theory allows for an analysis of market integration, achieved when short-sellers are allowed to deliver either asset.[9] We show that integration raises liquidity and welfare but not necessarily the price level.

This paper is closely related to Duffie's [1996] theory of repo specialness. In Duffie, short-sellers need to borrow an asset and sell it in the market, incurring an exogenous transaction cost. Assets that can be sold at a low cost are on special because they are in high demand by short-sellers. The main difference with Duffie is that instead of explaining specialness taking liquidity (transaction costs) as exogenous, we explain liquidity and specialness simultaneously. Thus, we can explain why liquidity and specialness might differ for otherwise identical assets. We can also analyze the effects of issue size, market integration, etc, taking into account the endogenous variation in liquidity. Krishnamurthy [2002] proposes a model building on Duffie [1996] that links the specialness premium to an exogenous liquidity premium. This link is also present in our model where the liquidity premium is endogenous.[10]

This paper builds on a series of papers by Duffie, Gârleanu and Pedersen, who are the first to introduce search in models of asset market equilibrium. Duffie, Gârleanu and Pedersen [2004] consider a model where investors seek to establish long positions, and Duffie, Gârleanu and Pedersen [2005] introduce dealers into that model. Duffie, Gârleanu and Pedersen [2002] introduce short-sellers into a different model where the spot market is Walrasian but the repo market operates through search. They show that specialness arises because of lenders' bargaining power, exactly as in this paper. Our focus differs in that we seek to explain differences in liquidity across assets. This leads us to extend their framework in several technically challenging directions. In particular, we consider a multi-asset model while they assume only one asset. We also introduce search in both the spot and the repo market because this is crucial for our explanation.[11] Vayanos and Wang [2004] and Weill [2004] develop multi-asset models with search, but no short-sellers. We show that

[8] For example, Amihud and Mendelson [1991] argue that a bond's effective supply decreases over time as the bond becomes "locked away" in institutional investors' portfolios.

[9] Bennett, Garbade and Kambhu [2000] argue that market integration can be achieved if on- and off-the-run bonds become standardized in terms of their maturity dates. For example, a two-year bond can be designed to mature on exactly the same date as a previously-issued five-year bond. The bonds can then be made "fungible," assigned the same CUSIP number, and be identical for delivery purposes.

[10] Empirical studies by Cornell and Shapiro [1989], Jordan and Jordan [1997], Buraschi and Menini [2002], Krishnamurthy [2002], Graveline and McBrady [2004], and Moulton [2004] show that on-the-run bond prices contain specialness premia consistent with Duffie [1996] and our model. We return to these studies in Section 7.

[11] Search in the spot market induces short-sellers to concentrate in one asset. Search in the repo market generates a positive lending fee, which is necessary to rule out the arbitrage strategy of shorting the on-the-run bond and buying the off-the-run.

in the presence of short-sellers differences in liquidity arise quite naturally. Moreover, price effects can be large because of the specialness premium.

This paper is related to the monetary-search literature building on Kiyotaki and Wright [1989] and Trejos and Wright [1995]. Aiyagari, Wallace and Wright [1996] provide an example of an economy in which fiat monies (intrinsically worthless and unbacked pieces of paper) endogenously differ in their price and liquidity. Wallace [2000] analyzes the relative liquidity of currency and dividend-paying assets in a model based on asset indivisibility. Our relative contribution is to compare dividend-paying assets as opposed to currency, and introduce short sales. The latter allows to examine whether price differences between otherwise identical assets can generate arbitrage.

This paper is also related to the literature on equilibrium asset pricing with transaction costs. (See, for example, Amihud and Mendelson [1986], Constantinides [1986], Aiyagari and Gertler [1991], Heaton and Lucas [1996], Vayanos [1998], Vayanos and Vila [1999], Huang [2003], and Lo, Mamaysky and Wang [2004].) Besides endogenizing the transaction costs, we add to that literature by introducing short-sales.

Pagano [1989] studies the concentration of liquidity across market venues. He shows that markets can coexist, but the outcome is generally dominated by concentrating trade in one market and shutting the other.[12] Our model differs because we consider concentration across assets rather than markets. In particular, there is no analogue of a market being shut.

Boudoukh and Whitelaw [1993] propose a theory of the on-the-run phenomenon in which liquidity is selected by the bond issuer. They show that the issuer can achieve price discrimination by imposing liquidity differences between otherwise identical bonds. This resembles our result that relative to market integration, the asymmetric (on-the-run) equilibrium can increase government revenue but reduce welfare.

The rest of this paper is organized as follows. Section 2 presents the model, and Section 3 considers the benchmark case where markets are Walrasian. Section 4 assumes that markets operate through search, and contains our main results. Section 5 extends the model to different asset supplies and market integration. Section 6 calibrates the model, Section 7 examines the model's empirical implications, and Section 8 concludes. All proofs are in the Appendix.

[12] See also Ellison and Fudenberg [2003] for a general analysis of the coexistence of markets, and Economides and Siow [1988] for a spatial model of market formation. See also Admati and Pfleiderer [1988] and Chowdhry and Nanda [1991] for models where trading is concentrated in a specific time or location because of asymmetric information.

2 Model

Time is continuous and goes from zero to infinity. There is a riskless asset with an exogenous return r, and two risky assets paying the same cash flow. Cash flow is described by the cumulative dividend process

$$dD_t = \delta dt + \sigma dB_t, \tag{1}$$

where δ and σ are positive constants, and B_t is a standard Brownian motion.[13] The risky assets can differ in their supply, and we denote by S_i the number of shares of asset $i \in \{1, 2\}$.

There is an infinite mass of infinitely-lived risk-averse agents who derive utility from the consumption of a numéraire good. All agents have the same CARA utility function,

$$-E\left[\int_0^\infty \exp\left(-\alpha c_t - \beta t\right)\, dt\right]. \tag{2}$$

Agents differ in their endowment streams. An agent can either receive an endowment that is positively correlated with dividends, or one that is negatively correlated, or one that is uncorrelated. The correlation between endowments and dividend give rise to hedging demands, inducing agents to trade. We refer to the agents with the negatively correlated endowment as high-valuation because they have a positive hedging demand. Likewise, we refer to the agents with the positively correlated endowment as low-valuation, and to those with the uncorrelated endowment as average-valuation.[14] Following Duffie, Gârleanu and Pedersen [2004], we assume that an agent receives the cumulative endowment process

$$de_t = \sigma_e \left[\rho_t dB_t + \sqrt{1 - \rho_t^2} dZ_t\right], \tag{3}$$

where σ_e is a positive constant and Z_t is a standard Brownian motion independent of B_t. The process ρ_t is the instantaneous correlation between the dividend process and the agent's endowment process. We set $\rho_t = -\overline{\rho} < 0$ for high-valuation agents, $\rho_t = \underline{\rho} > 0$ for low-valuation agents, and $\rho_t = 0$ for average-valuation agents. The processes (ρ_t, Z_t) are taken to be pairwise independent across agents.

[13] The process (1) is the continuous-time analog of *i.i.d.* cash flows. In a fixed-income setting, cash flows are deterministic and the uncertainty arises because of interest rates. Moreover, assets generally have a finite maturity rather than being infinitely lived. We abstract from these complications to keep the model tractable, but we believe that the basic intuitions are robust.

[14] The endowments can be interpreted as a position in a correlated market. For example, low-valuation agents could have long positions in corporate bonds or mortgage-backed securities, and seek to hedge them by shorting Treasuries. For a discussion of hedging demand in the Treasury market, see Dupont and Sack [1999].

There is a continuous flow $\overline{F}$ of average-valuation agents who switch to high valuation, and a continuous flow $\underline{F}$ who switch to low valuation. Conversely, high-valuation agents revert to average valuation with Poisson intensity $\overline{\kappa}$, and low-valuation agents do the same with Poisson intensity $\underline{\kappa}$. Thus, in a steady state, the measures of high- and low-valuation agents are $\overline{F}/\overline{\kappa}$ and $\underline{F}/\underline{\kappa}$, respectively. Given that the measure of average-valuation agents is infinite, an individual agent's switching intensity from average to high or low valuation is zero.

Agents can hold long, short, or no positions in any asset. Positions must be in multiples of a "round lot" that we normalize to one share. We are interested in steady-state equilibria where high-valuation agents are long one share or hold no position, low-valuation agents are short one share or hold no position, and average-valuation agents stay out of the market. In the following sections we show that such equilibria exist under appropriate parameter restrictions.

3 Walrasian Equilibrium

In this section we consider the benchmark case where markets are Walrasian, and show that both assets must trade at the same price. For notational simplicity, we set $A \equiv r\alpha$, $y \equiv A\sigma^2/2$, $\overline{x} \equiv A\overline{\rho}\sigma\sigma_e$, and $\underline{x} \equiv A\underline{\rho}\sigma\sigma_e$.

Proposition 1 *In a Walrasian equilibrium, both risky assets trade at the same price p. If*

$$\frac{\overline{F}}{\overline{\kappa}} > \sum_{i=1}^{2} S_i + \frac{\underline{F}}{\underline{\kappa}} \tag{4}$$

and

$$4y > \overline{x} + \underline{x} > 2y > \overline{x}, \tag{5}$$

then high-valuation agents either buy one share or stay out of the market, low-valuation agents short one share, average-valuation agents stay out of the market, and the price is

$$p = \frac{\delta + \overline{x} - y}{r}. \tag{6}$$

That both assets trade at the same price follows from no-arbitrage: since assets have the same cash flow and there are no trading frictions, an agent could make an infinite profit from a price

discrepancy. Note also that asset owners are lending their asset to short-sellers for a zero fee. (Repo transactions are introduced more explicitly in Section 4.1.) Indeed, if the fee were positive, all owners would be willing to lend, and the supply of lendable assets would always exceed the short-sellers' borrowing demand.

To explain the intuition for the rest of the proposition, we consider agents' portfolio-choice problem. An agent who holds z_t shares of the risky assets receives the instantaneous cash flow $z_t dD_t + de_t$. In Appendix A we show that the agent chooses z_t to maximize the mean-variance objective

$$E_t(z_t dD_t + de_t) - \frac{A}{2}\mathrm{Var}_t(z_t dD_t + de_t) - rpz_t dt,$$

where A is the (constant) coefficient of absolute risk aversion of the agent's value function. From Equations (1) and (3), this objective is equivalent to

$$\delta z_t - \frac{A}{2}\left(\sigma^2 z_t^2 + 2\rho_t \sigma \sigma_e z_t\right) - rpz_t \equiv C(\rho_t, z_t) - rpz_t,$$

where $C(\rho, z)$ is the incremental certainty equivalent of holding z shares relative to holding none. Using the definitions of y, $\overline{x}$, $\underline{x}$, we can write the certainty equivalent as $C(\overline{\rho}, z) = (\delta + \overline{x})z - yz^2$ for a high-valuation agent, $C(\underline{\rho}, z) = (\delta - \underline{x})z - yz^2$ for a low-valuation agent, and $C(0, z) \equiv \delta z - yz^2$ for an average-valuation agent. The parameter y measures the cost of bearing risk, and the parameters $\overline{x}$ and $\underline{x}$ measure the hedging benefits.

The aggregate asset supply is the sum of the supplies S_i, $i \in \{1, 2\}$, from the issuers, plus the supply generated by the short-sellers. Let's guess (and later verify) that low-valuation agents are the only short-sellers and short one share, in which case the latter supply is equal to their measure $\underline{F}/\underline{\kappa}$. Equation (4) then implies that the measure $\overline{F}/\overline{\kappa}$ of high-valuation agents exceeds the aggregate supply. Therefore, in equilibrium high-valuation agents must be indifferent between buying one share or none, and the price must equal $C(\overline{\rho}, 1)/r$, the present value (PV) of their certainty equivalent of one share.[15] Equation (5) ensures that our guess is verified, i.e., at the price $C(\overline{\rho}, 1)/r$, low-valuation agents find it optimal to short one share and average-valuation agents to stay out of the market.

[15] Intuitively, Equation (4) ensures that asset demand, generated by the high-valuation agents, exceeds asset supply. This implies that buyers are the "long" side of the market and bid the price up to their valuation. Equation (4) simplifies our analysis in several respects. For example, it ensures the existence of a parameter region in which short-sellers are the infra-marginal traders (Equation (13)).

4 Search Equilibrium

In this section we assume that markets operate through search and bilateral bargaining. There are two markets in our economy: the spot market for buying and selling, and the repo market where short-sellers can borrow an asset. We assume that both operate through search, although they can differ in the efficiency of the search process. In Section 4.1 we motivate our basic framework, and in Section 4.2 we describe agents' life-cycles and the search process. In Section 4.3 we determine the measures of the different agent types, in Section 4.4 we solve the agents' optimization problems, and in Section 4.5 we solve for the equilibrium. Throughout, we assume that asset supplies are identical, i.e., $S_1 = S_2 = S$, and allow for different supplies in Section 5.

4.1 Basic Framework

We adopt the standard search framework where agents are matched randomly over time in pairs. This framework captures some elements of the government-bond market: negotiations are mainly bilateral and locating counterparties can involve search. On the other hand, the framework is quite stylized. For example, it leaves open the question why agents cannot gather in a centralized marketplace.

In some sense, every price-formation model is stylized. For example, the Walrasian auction assumes multilateral trade, but the government-bond market operates mainly in a decentralized fashion through bilateral negotiations. Therefore, as long as search times are reasonably small, it is not obvious which model describes the market better. Furthermore, an explanation of the on-the-run phenomenon requires some sort of friction, which is absent from the Walrasian auction. Of course, the search framework is not the only way to introduce friction. A leading alternative is to assume asymmetric information about asset payoffs, but it is unclear what the asymmetries are in the government-bond market.[16] Perhaps other alternatives could be explored, but search offers a analytically tractable and parsimonious one.[17]

Before turning to our model of the repo market, we recall the mechanics of a repo transaction. In a repo transaction a lender turns his asset to a borrower in exchange for cash. At maturity the borrower returns the asset, and the lender returns the cash together with some previously-agreed

[16] See Admati and Pfleiderer [1992] for an asymmetric-information model where identical assets can differ in liquidity.

[17] A basic property of the search framework is that the time to execute a transaction decreases in the number of potential counterparties. For example, if short-sellers concentrate in one asset, this asset will have a larger buyer and seller pool, and be easier to trade. The basic mechanisms that we identify in this paper should carry to any model that shares this property.

interest-rate payment, called the "repo rate." Hence, a repo transaction is effectively a loan of cash collateralized by the asset. Treasury securities differ in their repo rates. Most of them share the same rate, called the "general collateral rate," which is the highest quoted repo rate and is close to the Fed Funds Rate. The specialness of an asset is defined as the difference between the general collateral rate and its repo rate. In our model, instead of assuming that the lender pays a repo rate to the borrower, we assume that the borrower pays a flow fee w to the lender. Hence, the "implied" repo rate is the difference $r - w/p$ between the risk-free rate and the lending fee per dollar, and the specialness is simply w/p.

Finally, we impose from now on a simplifying restriction on agents' portfolios. We assume that agents can either hold a long position of one share, or a short position of one share, or no position. Precluding long and short positions of multiple shares is relatively innocuous. Indeed, Section 3 shows that agents can choose to limit themselves to one share because of risk aversion. The less innocuous part of the constraint is to preclude arbitrage portfolios of offsetting long and short positions. To impose the discipline of no-arbitrage on the market, we introduce an additional agent group, the "arbitrageurs." These are average-valuation agents who never switch to high or low valuation, and who can hold either one of the three portfolios above or an arbitrage portfolio that is one share long and one short. We assume that arbitrageurs have infinite measure so that they can take an unlimited collective position.[18]

4.2 Agents' Life-Cycles and the Search Process

We look for equilibria in which portfolio decisions resemble those in the Walrasian case, namely, high-valuation agents seek to buy one share of an asset, low-valuation agents seek to short one share, and average-valuation agents (including arbitrageurs) stay out of the market. Agents' life-cycles in these equilibria are roughly as follows. High-valuation agents buy the asset, possibly lend it to a short-seller, and then sell it when switching to average valuation. Low-valuation agents borrow the asset, sell it, and then buy it back when switching to average valuation.

A full description of life-cycles must include some additional outcomes. For example, a high-valuation agent could switch to average valuation before buying the asset. Alternatively, the switch could occur while the asset is on loan, in which case we must specify how the loan is terminated. To represent the full range of outcomes, we use the flow diagram of Figure 1.

[18] Alternatively, we could assume that all agents can hold the arbitrage portfolio, and do away with the arbitrageurs. We return to this issue in Footnote 29.

The top half of the diagram refers to a high-valuation agent. The agent is initially a high-valuation buyer $\overline{b}$, seeking a seller of either asset in the spot market. If he reverts to average valuation before meeting a seller, he exits the market. Otherwise, if he meets a seller of asset $i \in \{1,2\}$, he buys the asset. He then becomes a lender $\overline{\ell}i$ of asset i in the repo market, seeking a borrower. If he reverts to average valuation before meeting a borrower, he exits the repo market and becomes a seller $\overline{s}i$ of asset i in the spot market. Otherwise, if he meets a borrower and there are gains from trade, he lends the asset.

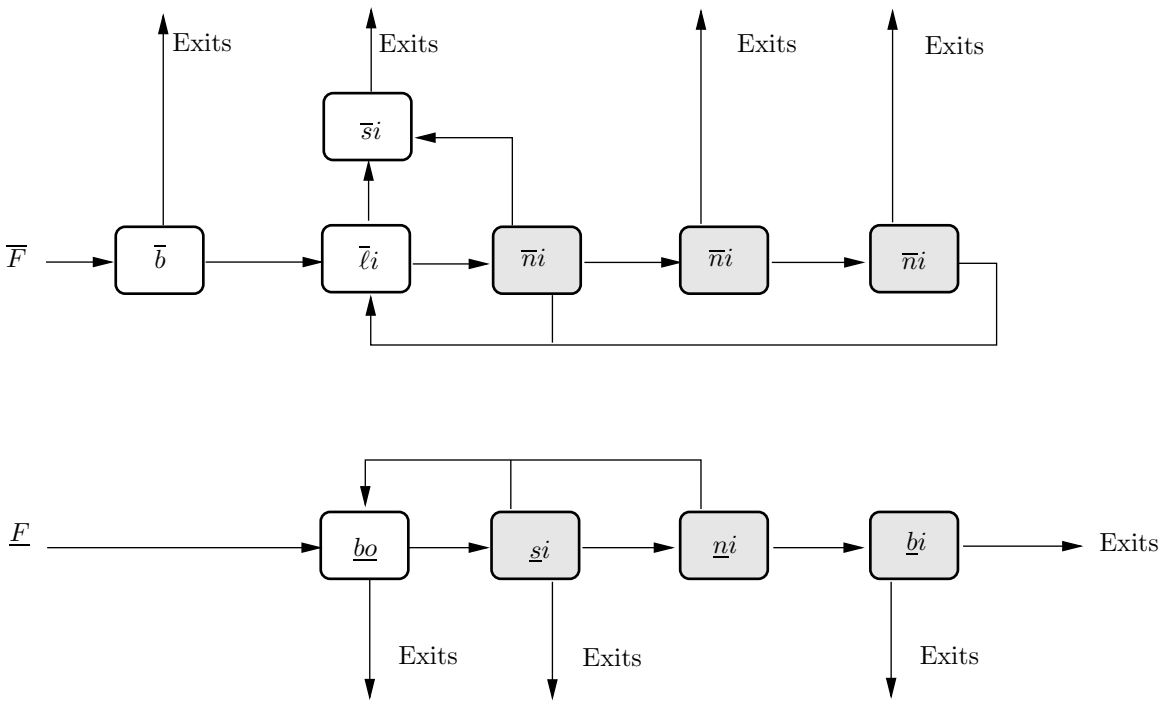
To describe the subsequent outcomes, we must consider the low-valuation agent who borrows the asset, and we turn to the bottom half of the diagram. Initially, a low-valuation agent is a borrower $\underline{bo}$, seeking a lender in the repo market. If she reverts to average valuation before meeting a lender, she exits the market. Otherwise, if she meets a lender of asset i and there are gains from trade, she borrows the asset.

After the loan is agreed, the borrower $\underline{bo}$ becomes a low-valuation seller $\underline{s}i$ of asset i, and the lender $\overline{\ell}i$ becomes a high-valuation non-searcher $\overline{n}i$, represented by the box above $\underline{s}i$. If agent $\underline{s}i$ reverts to average valuation before meeting a buyer, she delivers the asset to agent $\overline{n}i$ and exits the market, while agent $\overline{n}i$ returns to the lender pool $\overline{\ell}i$. If instead it is agent $\overline{n}i$ who reverts to average valuation, agent $\underline{s}i$ delivers the asset and returns to the borrower pool $\underline{bo}$, while agent $\overline{n}i$ becomes a seller $\overline{s}i$. Otherwise, if agent $\underline{s}i$ meets a buyer, she sells the asset and becomes a low-valuation non-searcher $\underline{n}i$. Agent $\overline{n}i$ then moves to the box above $\underline{n}i$. If agent $\underline{n}i$ reverts to average valuation after that time, she becomes a buyer $\underline{b}i$ of asset i, and agent $\overline{n}i$ moves to the box above $\underline{b}i$. Upon meeting a seller, agent $\underline{b}i$ buys and delivers the asset, while agent $\overline{n}i$ returns to the lender pool $\overline{\ell}i$. If agent $\overline{n}i$ reverts to average valuation while his counterparty is $\underline{n}i$ or $\underline{b}i$, then instant delivery is impossible because of search. In that event, we assume that agent $\overline{n}i$ seizes some cash collateral previously posted by the low-valuation agent, and exits the market.[19]

We refer to the different states in agents' life-cycles as "types." The types, summarized in the table below Figure 1, are $\overline{b}$ and $\{\overline{\ell}i, \overline{n}i, \overline{s}i\}_{i\in\{1,2\}}$ for the high-valuation agents, and $\underline{bo}$ and $\{\underline{s}i, \underline{n}i, \underline{b}i\}_{i\in\{1,2\}}$ for the low-valuation agents. We denote by $\mathcal{T}$ the set of types, and by μ_τ the measure of investors of type $\tau \in \mathcal{T}$. Finally, we denote by bi the group of all buyers of asset i (both

[19]This assumption is for simplicity. An alternative assumption is that the low-valuation agent can search for the asset under a late-delivery penalty, but this would complicate the model without changing the basic intuitions.

In Appendix D we show that because collateral acts as a transfer, its specific value does not affect any equilibrium variable except the price of the repo contract: high-valuation agents accept to lend their asset for a lower fee if they can seize more collateral. To downplay this effect, we set the collateral equal to the utility of a seller $\overline{s}i$. This ensures that upon reverting to average valuation, agent $\overline{n}i$ is equally well off when receiving the asset (thus becoming a seller $\overline{s}i$) or the cash collateral.



High-valuation investors (top half)	
High-valuation buyer	$\overline{b}$
Lender of asset $i \in \{1,2\}$	$\overline{\ell}i$
High-valuation seller of asset $i \in \{1,2\}$	$\overline{s}i$
High-valuation non searcher	$\overline{n}i$

Low-valuation investors (bottom half)	
Borrower	$\underline{bo}$
Low-valuation seller of asset $i \in \{1,2\}$	$\underline{s}i$
Low-valuation non searcher	$\underline{n}i$
Low-valuation buyer of asset $i \in \{1,2\}$	$\underline{b}i$

In the diagram, boxes represent investors' types and arrows transitions between types.
The shaded boxes represent types who are matched through a repo transaction.
Our notation for investors' types is summarized in the table.

Figure 1: Flow Diagram.

high- and low-valuation), and by si the group of all sellers. The measures μ_{bi} and μ_{si} of these groups are

$$\mu_{bi} = \mu_{\overline{b}} + \mu_{\underline{b}i} \tag{7}$$

$$\mu_{si} = \mu_{\overline{s}i} + \mu_{\underline{s}i}. \tag{8}$$

In each market agents are matched randomly over time in pairs. We assume that an agent

establishes contact with other agents at Poisson arrival times with fixed intensity. Moreover, there is random matching in that conditional on establishing a contact, all agents are "equally likely" to be contacted. Thus, an agent meets members of a given group with Poisson intensity proportional to that group's measure. For example, a buyer in the spot market meets sellers of asset i with Poisson intensity $\lambda\mu_{si}$, where λ is a parameter measuring the efficiency of spot-market search. Therefore, the Law of Large Numbers (see Duffie and Sun [2004]) implies that meetings between buyers and sellers of asset i occur at a deterministic rate $\lambda\mu_{bi}\mu_{si}$. Likewise, meetings between borrowers and lenders of asset i occur at a deterministic rate $\nu\mu_{\underline{bo}}\mu_{\overline{\ell}i}$, where ν measures the efficiency of repo-market search. If in equilibrium low-valuation agents decide to borrow only one asset, some of the borrower-lender meetings do not result in a trade. To account for this, we define the endogenous variable ν_i by $\nu_i = \nu$ if low-valuation agents borrow asset i, and $\nu_i = 0$ otherwise.

When two agents meet, they bargain over the terms of trade. Bargaining in the spot market is over the price p_i of asset i, and in the repo market is over the flow fee w_i that the borrower must pay to the lender of asset i over the life of the loan. We assume that bargaining takes place according to a simple game where the two agents make simultaneous offers. If the offers generate a set of mutually acceptable prices, trade occurs at the mid-point of that set. Otherwise, the meeting ends and the agents return to the search pool.

4.3 Demographics

We next derive a set of equations that determine the steady-state measures of the different agent types. Market clearing requires that assets are held by lenders or sellers, i.e.,

$$\mu_{\overline{\ell}i} + \mu_{si} = S_i, \tag{9}$$

for $i \in \{1, 2\}$. Moreover, the measure $\mu_{\overline{n}i}$ of high-valuation agents engaged in a borrower-lender match must equal the measure of low-valuation agents, i.e.,

$$\mu_{\overline{n}i} = \mu_{\underline{s}i} + \mu_{\underline{n}i} + \mu_{\underline{b}i}, \tag{10}$$

for $i \in \{1, 2\}$. The remaining equations follow from the requirement that the inflow into an agent type must equal the outflow. Consider, for example, the type $\overline{b}$ of high-valuation buyers. The inflow into this type is $\overline{F}$ because of the new entrants. The outflow is the sum of $\overline{\kappa}\mu_{\overline{b}}$ because some high-valuation buyers revert to average valuation and exit the market, and $\sum_{i=1}^{2} \lambda\mu_{si}\mu_{\overline{b}}$ because

some high-valuation buyers meet sellers of assets 1 or 2 and buy. Therefore,

$$\overline{F} = \overline{\kappa}\mu_{\overline{b}} + \sum_{i=1}^{2} \lambda \mu_{si} \mu_{\overline{b}}. \tag{11}$$

In Appendix B we derive the remaining inflow-outflow equations and show that the resulting system has a unique solution.

4.4 Optimization

Agents optimize over their consumption flow and risky-asset portfolio. We solve the optimization problem in two steps: (i) take the portfolio decision as given and solve for optimal consumption, thus computing an "interim" value function, and (ii) determine the portfolio decision that maximizes this value function. We characterize the value function in Appendix C, and leave portfolio optimization to Appendix E where we compute the full equilibrium. In this section we present an intuitive characterization of the value function corresponding to agents' equilibrium portfolio decisions, i.e., the life-cycles of Section 4.2.

An agent's value function takes the form

$$-\frac{1}{r}\exp\left[-A\left(W_t + V_\tau\right) + \frac{r - \beta + A^2\sigma_e^2/2}{r}\right],$$

where W_t is the investment in the riskless asset, and V_τ is a constant that depends on the agent's type and to which we refer as the agent's "utility." The set of utilities solves a system of nonlinear equations. The equations, however, become linear and can be solved in closed form in an interesting special case. This is when the coefficient of absolute risk aversion A goes to zero, holding the parameters y, $\overline{x}$, and $\underline{x}$ constant.[20] Because these parameters measure the cost and hedging benefit of bearing risk, risk considerations matter even in the limit. It is, however, only the risk of the dividend process that matters, and not the risk associated to the matching process in the search market. Agents are effectively risk-neutral relative to the latter, and this is what makes the equations linear. From now on, we focus on the limit case because it captures the key economic intuitions while also generating simpler equations.

[20] Recall from Section 3 that these parameters are defined by $y \equiv A\sigma^2/2$, $\overline{x} \equiv A\overline{\rho}\sigma\sigma_e$, and $\underline{x} \equiv A\underline{\rho}\sigma\sigma_e$. Therefore, when A goes to zero, the variances σ and σ_e must go to infinity. Note that the certainty equivalent $C(\rho, z)$ is unaffected when taking the limit because its dependence on A, σ, and σ_e is only through y, $\overline{x}$, and $\underline{x}$.

The equations take a form which is standard in the search literature. This is that the flow value rV_τ of being type τ is derived from the flow benefits accruing to that type (dividends and lending fees) plus the transitions to other types. For a high-valuation buyer $\overline{b}$, for example, the equation is

$$rV_{\overline{b}} = -\overline{\kappa}V_{\overline{b}} + \sum_{i=1}^{2} \lambda\mu_{si}(V_{\overline{\ell}i} - p_i - V_{\overline{b}}), \tag{12}$$

because the flow benefits are zero and the transitions are (i) revert to average valuation at rate $\overline{\kappa}$ and exit the market (utility zero and net utility $-V_{\overline{b}}$), and (ii) meet a seller of asset $i \in \{1, 2\}$ at rate $\lambda\mu_{si}$, buy at price p_i, and become a lender $\overline{\ell}i$ (utility $V_{\overline{\ell}i}$ and net utility $V_{\overline{\ell}i} - p_i - V_{\overline{b}}$).

In Appendix D we derive the remaining equations. These must be solved together with the equations for the price and the lending fee. The price is determined by bargaining between buyers and sellers. There are two types of buyers, $\overline{b}$ and $\underline{b}i$, and two types of sellers, $\overline{s}i$ and $\underline{s}i$. Type $\overline{b}$ has reservation value $\Delta_{\overline{b}} \equiv V_{\overline{\ell}i} - V_{\overline{b}}$ because after buying asset i he becomes a lender with utility $V_{\overline{\ell}i}$. Type $\underline{b}i$ has reservation value $\Delta_{\underline{b}i} \equiv -V_{\underline{b}i}$ because after buying she delivers the asset and exits the market. Likewise, the sellers' reservation values are $\Delta_{\overline{s}i} \equiv V_{\overline{s}i}$ and $\Delta_{\underline{s}i} \equiv V_{\underline{s}i} - V_{\underline{n}i}$. Because type $\overline{b}$ receives a hedging benefit from holding the asset while type $\overline{s}i$ does not, reservation values satisfy $\Delta_{\overline{b}} > \Delta_{\overline{s}i}$. They also satisfy $\Delta_{\underline{b}i} > \Delta_{\underline{s}i}$ because type $\underline{s}i$ receives a hedging benefit from holding a short position while type $\underline{b}i$ does not. To complete the ranking, we assume that short-sellers are the infra-marginal traders, both as sellers and as buyers, i.e.,

$$\Delta_{\underline{b}i} > \Delta_{\overline{b}} > \Delta_{\overline{s}i} > \Delta_{\underline{s}i}. \tag{13}$$

This assumption makes the analysis more transparent because it ensures that the marginal traders are comparable across assets, even in equilibria where short-selling is concentrated on one asset. In Section 4.5 we show that Equation (13) is satisfied under appropriate restrictions on exogenous parameters.

We focus on simple equilibria of the bargaining game in which all buyers and sellers make the same offer p_i. This offer must be in $[\Delta_{\overline{s}i}, \Delta_{\overline{b}}]$ to ensure that all traders realize a non-negative surplus. Given the buyers' strategy, asking p_i is optimal for a seller - a higher ask would preclude trading while a lower ask would lower the transaction price. Likewise, given the sellers' strategy, bidding p_i is optimal for a buyer. Obviously any $p_i \in [\Delta_{\overline{s}i}, \Delta_{\overline{b}}]$ is an equilibrium. We do not select

among these, but instead treat the buyers' "bargaining power" ϕ defined by

$$p_i = \phi \Delta_{\overline{s}i} + (1 - \phi)\Delta_{\overline{b}}, \tag{14}$$

as exogenous. The bargaining power ϕ is equal to the fraction of the overall surplus $\Delta_{\overline{b}} - \Delta_{\overline{s}i}$ that the marginal buyer $\overline{b}$ can extract.

The lending fee is determined by bargaining between borrowers and lenders. We compute it in Appendix D as a function of the surplus Σ_i associated to a borrower-lender match, and the fraction θ of that surplus that a lender can extract.

4.5 Equilibrium

An equilibrium is characterized by

(i) Measures μ_τ for all agent types $\tau \in \mathcal{T}$.

(ii) Utilities V_τ for all agent types $\tau \in \mathcal{T}$.

(iii) Prices and lending fees (p_i, w_i) for $i \in \{1, 2\}$.

(iv) Short-selling decisions $\nu_i \in \{0, \nu\}$ for $i \in \{1, 2\}$.

These variables are a solution to the following fixed-point problem. The measures are determined from the nonlinear system of Equations (9)-(11) and (B.1)-(B.6), as a function of the short-selling decisions. The utilities, prices, and lending fees are determined from the linear system of equations (12) and (D.1)-(D.11), as a function of the measures and short-selling decisions. Finally, the short-selling decisions are determined as a function of the utilities from

$$\nu_i = \nu \Leftrightarrow \Sigma_i \geq 0, \tag{15}$$

i.e., agents short-sell asset i if the surplus Σ_i associated to a borrower-lender match is positive.

A solution to the fixed-point problem is an equilibrium if it satisfies two additional requirements. First, the conjectured portfolio decisions must be optimal, i.e., high- and low-valuation agents must adopt the life-cycles of Section 4.2, and average-valuation agents (including arbitrageurs) must hold no position. Second, the buyers' and sellers' reservation values must be ordered as in Equation (13).

We are interested in two types of equilibria: a symmetric one where low-valuation agents short-sell both assets, i.e., $\nu_1 = \nu_2 = \nu$, and an asymmetric one where short-selling is concentrated on one asset only, say asset 1, i.e., $\nu_1 = \nu$ and $\nu_2 = 0$. Computing these equilibria can, in general, be done only numerically. Fortunately, however, closed-form solutions can be derived when search frictions are small, i.e., λ and ν are large.[21] In the remainder of this section we focus on this case, emphasizing the intuitions gained by the closed-form solutions. We complement our asymptotic analysis with a numerical calibration in Section 6.

When search frictions are small, the measure of agents in the "short" side of a market goes to zero. The short side in the repo market are the borrowers because they enter the market at a flow rate, while the lenders are the asset-holders and constitute a stock. The short side in the spot market depends on the comparison between the asset demand, generated by the high-valuation agents, and the asset supply, generated by the issuers and the short-sellers. As in the Walrasian case, we assume that demand exceeds supply, i.e.,

$$\frac{\overline{F}}{\overline{\kappa}} > 2S + \frac{\underline{F}}{\underline{\kappa}}. \tag{16}$$

Under this condition, the short side in the spot market are the sellers.

4.5.1 Symmetric Equilibrium

Proposition 2 *Assume Equation (16),*

$$\frac{\underline{x} + \frac{\underline{\kappa}}{r+\overline{\kappa}+g_s}\overline{x}}{1 + \frac{\underline{\kappa}}{r+\overline{\kappa}+g_s}} > 2y > \overline{x}, \tag{17}$$

and $\phi, \theta \neq 1$, where g_s is defined by Equation (B.28) of Appendix D. Then, for large λ and ν, there exists a symmetric equilibrium in which prices, lending fees, and types' measures are identical across assets.

In the proof of Proposition 2 we determine the asymptotic behavior of the equilibrium. We confirm that the measures of sellers and borrowers, who are the short side in their respective markets, go to zero, while the measures of buyers and lenders go to positive limits. In particular, for each asset $i \in \{1, 2\}$, the measure of lenders converges to the asset supply S, and the measure

[21] More precisely, we assume that λ and ν go to ∞, holding the ratio $n \equiv \nu/\lambda$ constant. When taking this limit, we will say that a variable Z is asymptotically equal to $z_1/\lambda + z_2/\nu$, if $Z = z_1/\lambda + z_2/(n\lambda) + o(1/\lambda)$.

of buyers to a limit m_b. On the other hand, the measure of sellers is asymptotically equal to g_s/λ, and that of borrowers to $g_{\underline{bo}}/\nu$, for two constants g_s and $g_{\underline{bo}}$. The asymptotic behavior of the price and the lending fee is described in the following proposition.

Proposition 3 *In the symmetric equilibrium of Proposition 2, the price of each asset $i \in \{1,2\}$ is asymptotically equal to*

$$p_i = \frac{\delta + \overline{x} - y}{r} - \frac{\overline{\kappa}}{\lambda m_b}\frac{\overline{x}}{r} - \frac{\phi(r+\overline{\kappa}+2g_s)}{\lambda(1-\phi)m_b}\frac{\overline{x}}{r} + \frac{g_{\underline{bo}}}{r+\overline{\kappa}+\underline{\kappa}\frac{g_s}{r+\overline{\kappa}+\underline{\kappa}+g_s}+g_{\underline{bo}}}\frac{w_i}{r}, \tag{18}$$

and the lending fee is asymptotically equal to

$$w_i = \theta\left(r+\overline{\kappa}+\underline{\kappa}\frac{g_s}{r+\overline{\kappa}+\underline{\kappa}+g_s}+g_{\underline{bo}}\right)\Sigma_i, \tag{19}$$

where

$$\Sigma_i = \frac{\underline{x} - \frac{r+\overline{\kappa}+\underline{\kappa}+g_s}{r+\overline{\kappa}+g_s}(2y-\overline{x})}{2\nu(1-\theta)S}. \tag{20}$$

The price is the sum of four terms. The first term, $(\delta+\overline{x}-y)/r$, is the limit to which the price converges when search frictions go to zero. Not surprisingly, this limit is the Walrasian price of Proposition 1. Recall that the Walrasian price is the PV of the high-valuation agents' certainty equivalent of one share. High-valuation agents bid the price up to their valuation because they are the long side in the market.

The remaining terms in Equation (18) are adjustments to the Walrasian price due to search frictions. The second term is a liquidity discount, arising because high-valuation agents incur a search cost when needing to sell the asset. This cost reduces their valuation and lowers the asset price. The liquidity discount decreases in the measure of buyers (m_b in the limit) because this reduces the time to sell the asset, and increases in the rate $\overline{\kappa}$ of reversion to average valuation because this reduces the investment horizon. Interpreting the search cost as a transaction cost, the liquidity discount is in the spirit of Amihud and Mendelson [1986].[22]

[22]Consistent with Amihud and Mendelson, the liquidity discount $\overline{\kappa}\overline{x}/(\lambda m_b r)$ is the PV of transaction costs incurred by a sequence of marginal buyers. Indeed, a high-valuation investor (the marginal buyer) reverts to average valuation at rate $\overline{\kappa}$. He then incurs an opportunity cost $\overline{x}$ of holding the asset, since he does not realize the hedging benefit, until he meets a new buyer at rate λm_b.

The third term is a discount arising because high-valuation agents have bargaining power in the search market and can extract some surplus from the sellers. This "bargaining" discount is present only when the buyers' bargaining power ϕ is non-zero.

The last term is a specialness premium, arising because high-valuation agents can earn a fee by lending the asset in the repo market. This fee is an additional cash flow derived from the asset and raises its price. The specialness premium is the PV of the asset's expected lending revenue, but is smaller that the PV w_i/r of a continuous stream of the lending fee. This is because lenders must search for borrowers and cannot ensure that their asset is on loan continuously. In fact, the time to meet a borrower does not converge to zero when search frictions become small. For small frictions, the flow of borrowers who enter into the market are matched almost instantly with lenders. Because, however, lenders are in positive measure, the meeting time from any given lender's viewpoint is finite.[23]

The lending fee arises because of the lenders' bargaining power θ in the repo market, exactly as in Duffie, Gârleanu and Pedersen [2002]. When the bargaining power is non-zero, the lenders can extract some of the short-selling surplus Σ_i from the borrowers. Of course, when search frictions become small, lenders can be contacted almost instantly, and competition among them drives the fee down to zero.

The short-selling surplus Σ_i increases in the hedging benefit $\underline{x}$ of the low-valuation agents. It also increases in g_s, which is the Poisson intensity at which sellers can be contacted in the limit.[24] The easier the sellers are to contact, the more attractive a short-sale becomes to a low-valuation agent because it is easier to buy the asset back.

4.5.2 Asymmetric Equilibrium

Proposition 4 *Assume Equations (16), (17), $\phi \neq 1$, $\theta \neq 0, 1$, and $\nu/\lambda \in (n_1, n_2)$ for two positive constants n_1, n_2. Then, for large λ and ν, there exists an asymmetric equilibrium where short-selling is concentrated on asset 1.*

[23] Formally, the measure of borrowers is asymptotically equal to $g_{\underline{bo}}/\nu$, and thus the Poisson intensity $\nu\mu_{\underline{bo}}$ at which borrowers can be contacted converges to $g_{\underline{bo}}$.

[24] The Poisson intensity at which sellers can be contacted is $\lambda\mu_s$, and converges to g_s because the measure of sellers is asymptotically equal to g_s/λ. The surplus is increasing in g_s because Equation (17) requires that $2y > \overline{x}$. This inequality ensures that upon reverting to average valuation, a short-seller prefers to buy the asset back rather than keeping the short position.

The surplus Σ_i is positive because of the left-hand-side inequality in Equation (17). In fact, this inequality is stronger than $\Sigma_i > 0$ because it ensures that low-valuation agents are not only willing to short-sell, but are also the infra-marginal, i.e., the more eager, sellers.

Taken together, Propositions 2 and 4 imply that there is a parameter range for which a symmetric and an asymmetric equilibrium coexist. In the asymmetric equilibrium, low-valuation agents short-sell only asset 1, declining any opportunities to borrow asset 2. This occurs because of search externalities. The more agents short-sell asset 1, the greater the asset's seller pool becomes. The asset's buyer pool also increases because of the short-sellers who need to buy the asset back. This makes asset 1 easier to trade, attracting, in turn, more short-sellers.

While the general notion of search externalities is well-understood, its application to the on-the-run phenomenon is subtle. Absent the short-sellers, search externalities would not operate. Indeed, the only agents choosing between the two assets would be the high-valuation buyers. While these agents value an asset with a larger buyer pool (because they eventually turn into sellers), prices would adjust so that in equilibrium high-valuation agents hold both assets. Therefore, the assets would have a common buyer pool, and be identical from a buyer's viewpoint.[25]

Search externalities would not operate even with short-sellers, if these were allowed to deliver any asset and not necessarily the one they borrowed. Indeed, the assets would have a common buyer pool, consisting of the high-valuation buyers and the short-sellers who need to deliver. Therefore, both assets would be equally attractive to short-sell: equally easy to sell because of the common buyer pool, and equally easy to deliver because one can be substituted for the other.

Summarizing, search externalities can operate only because of the combination of short-sellers and the constraint that these can deliver only the asset they borrowed. This constraint "locks" short-sellers into buying one asset, thus generating differences in the assets' buyer pools. It also implies that that the size of an asset's seller pool matters: a short-seller finds it more valuable to borrow an asset with a larger seller pool because that asset can be delivered more easily.

In the proof of Proposition 4, we determine the asymptotic behavior of the equilibrium. We show that for each asset i, the measure of lenders converges to the asset supply S, and the measure of buyers to a limit $\hat{m}_{bi}$ such that $\hat{m}_{b1} > \hat{m}_{b2}$. On the other hand, the measure of sellers is asymptotically equal to $\hat{g}_{si}/\lambda$, and that of borrowers to $\hat{g}_{\underline{bo}}/\nu$, for constants $\hat{g}_{s1} > \hat{g}_{s2}$ and $\hat{g}_{\underline{bo}}$. We return to these constants in Section 4.5.3, where we compare the symmetric and the asymmetric equilibrium. The asymptotic behavior of the price and the lending fee is in the following proposition.

Proposition 5 *In the asymmetric equilibrium of Proposition 4, asset prices are asymptotically*

[25]The proof that all assets would have a common buyer pool, and the same price, in the absence of short-sellers is available upon request. Note that the argument ruling out concentration of buyers does not apply to short-sellers. Indeed, while equilibrium requires that some agents hold long positions in each asset, there is no such requirement for short positions.

equal to

$$p_1 = \frac{\delta + \overline{x} - y}{r} - \frac{\overline{\kappa}}{\lambda \hat{m}_{b1}} \frac{\overline{x}}{r} - \frac{\phi}{\lambda(1-\phi)} \left[\frac{r + \overline{\kappa} + \hat{g}_{s1}}{\hat{m}_{b1}} + \frac{\hat{g}_{s2}}{\hat{m}_{b2}} \right] \frac{\overline{x}}{r} + \frac{\hat{g}_{b\underline{o}}}{r + \overline{\kappa} + \underline{\kappa} \frac{\hat{g}_{s1}}{r + \overline{\kappa} + \underline{\kappa} + \hat{g}_{s1}}} + \hat{g}_{b\underline{o}} \frac{w_1}{r} \quad (21)$$

and

$$p_2 = \frac{\delta + \overline{x} - y}{r} - \frac{\overline{\kappa}}{\lambda \hat{m}_{b2}} \frac{\overline{x}}{r} - \frac{\phi}{\lambda(1-\phi)} \left[\frac{r + \overline{\kappa} + \hat{g}_{s2}}{\hat{m}_{b2}} + \frac{\hat{g}_{s1}}{\hat{m}_{b1}} \right] \frac{\overline{x}}{r}. \quad (22)$$

The lending fee for asset 1 is asymptotically equal to

$$w_1 = \theta \left(r + \overline{\kappa} + \underline{\kappa} \frac{\hat{g}_{s1}}{r + \overline{\kappa} + \underline{\kappa} + \hat{g}_{s1}} + \hat{g}_{b\underline{o}} \right) \Sigma_1, \quad (23)$$

where

$$\Sigma_1 = \frac{\underline{x} - \frac{r + \overline{\kappa} + \underline{\kappa} + \hat{g}_{s1}}{r + \overline{\kappa} + \hat{g}_{s1}} (2y - \overline{x})}{\nu(1-\theta)S}. \quad (24)$$

An immediate consequence of Proposition 5 is that the price of asset 1 exceeds that of asset 2. This is because of three effects working in the same direction. First, the liquidity discount is smaller for asset 1 because this asset has a larger buyer pool, i.e., $\hat{m}_{b1} > \hat{m}_{b2}$. Second, the bargaining discount is smaller for asset 1 because the larger buyer pool implies more outside options for the sellers.[26] Finally, asset 1 carries a specialness premium because unlike asset 2, it can be lent to short-sellers.

Our model rationalizes the apparent paradox that off-the-run bonds are generally viewed as "scarce" and hard to locate, while at the same time being cheaper than on-the-run bonds. We show that off-the-run bonds are indeed scarce from the viewpoint of short-sellers seeking to buy and deliver them. At the same time, they are cheaper than on-the-run bonds because the marginal buyers are the agents seeking to establish long positions. These agents value the superior liquidity of the on-the-run bonds and the ability to lend the bonds in the repo market.

Since asset prices differ in the asymmetric equilibrium, a natural question is whether there exists a profitable arbitrage. By construction, an arbitrage cannot exist in our model because it would be eliminated by the group of arbitrageurs. The question is instead why arbitrageurs choose to hold no position even though asset prices differ.

[26] This logic does not apply to buyers because the marginal buyers are the high-valuation agents who are not limited to the seller pool of a specific asset.

Since asset 1 is more expensive than asset 2, an arbitrageur could buy asset 2 and short asset 1. The arbitrageur would, however, have to pay the lending fee for asset 1. Therefore, the strategy is unprofitable if

$$p_1 - p_2 < \frac{w_1}{r}, \tag{25}$$

i.e., the price differential between the two assets does not exceed the PV of the lending fee.[27] In equilibrium, however, the lending fee affects not only the cost of the arbitrage, but also the benefit: it raises the price differential through the specialness premium. To examine whether Equation (25) is satisfied, we thus need to substitute the equilibrium values of p_1 and p_2 from Proposition 5:

$$\frac{(\phi r + \overline{\kappa})}{\lambda(1-\phi)}\left[\frac{1}{\hat{m}_{b2}} - \frac{1}{\hat{m}_{b1}}\right]\frac{\overline{x}}{r} + \frac{\hat{g}_{\underline{bo}}}{r + \overline{\kappa} + \underline{\kappa}\frac{\hat{g}_{s1}}{r+\overline{\kappa}+\underline{\kappa}+\hat{g}_{s1}} + \hat{g}_{\underline{bo}}}\frac{w_1}{r} < \frac{w_1}{r}.$$

The first term on the left-hand side reflects asset 1's lower liquidity and bargaining discounts relative to asset 2, and we refer to it as asset 1's liquidity premium. By buying asset 2 and shorting asset 1, an arbitrageur capitalizes on this premium. The arbitrageur also capitalizes on the specialness premium, which is the second term on the left-hand side. Crucially, however, the specialness premium is only a fraction of the cost w_1/r of the arbitrage because lenders cannot ensure that their asset is on loan continuously (as emphasized in Section 4.5.1). Thus, Equation (25) is satisfied when the lending fee is large enough.[28]

An arbitrageur could follow the opposite strategy of buying asset 1 and shorting asset 2. In the proof of Proposition 4 we show that this strategy is unprofitable if

$$\frac{\hat{g}_{\underline{bo}}}{r + \underline{\kappa}\frac{\hat{g}_{s1}}{r+\underline{\kappa}+g_{s1}} + \hat{g}_{\underline{bo}}}\frac{w_1}{r} \leq p_1 - p_2. \tag{26}$$

The left-hand side is the arbitrageur's fee income from lending asset 1 in the repo market. This exceeds the specialness premium (included in $p_1 - p_2$) because the arbitrageur can hold asset 1

[27] In the proof of Proposition 4 we show that the strategy is unprofitable under the weaker condition $p_1 - p_2 < w_1/r + \xi$, for some some transaction cost ξ of establishing the arbitrage position: because trading opportunities arrive one at a time in a Poisson manner, it is not possible to set up the two legs of the position simultaneously, and this generates a cost of being unhedged for some time period.

[28] Our analysis has an interesting similarity to Krishnamurthy [2002], who assumes that $p_1 - p_2 = v + zw_1/r$, where v is a "liquidity benefit" of on-the-run bonds, and $z < 1$ is the extent to which bond owners can exploit the specialness premium. In our setting, v is the liquidity premium and z is determined by the lenders' search times.

One might argue that because of same-day settlement in the repo market, some sophisticated investors can manage to lend their asset almost continuously, i.e., $z \approx 1$. We conjecture that in a model with heterogenous lenders, the less sophisticated ones would have a lower reservation value for owning the asset and hence could be the "marginal buyers" in the spot market. The parameter z could then be significantly different than one, reflecting marginal buyers' inferior lending ability.

forever, thus being a better lender than a sequence of high-valuation agents. Because, however, the arbitrageur loses on the liquidity premium (the remaining part of $p_1 - p_2$), Equation (26) is satisfied when the lending fee is small enough. In the proof of Proposition 4 we show that Equations (25) and (26) are jointly satisfied when the ratio ν/λ of relative frictions in the spot and repo markets is in some interval (n_1, n_2).[29]

4.5.3 Comparison of the Symmetric and the Asymmetric Equilibrium

We next compare the equilibria of Propositions 2 and 4.

Proposition 6 *In the asymmetric equilibrium:*

(i) There are more buyers and sellers of asset 1 than in the symmetric equilibrium.

(ii) There are fewer buyers and sellers of asset 2 than in the symmetric equilibrium.

(iii) The lending fee of asset 1 is higher than in the symmetric equilibrium.

(iv) The prices of the two assets straddle the symmetric-equilibrium price when $\phi = 0$. For other values of ϕ (e.g., 1/2), both prices can exceed the symmetric-equilibrium price.

(v) Social welfare is higher than in the symmetric equilibrium.

Since in the asymmetric equilibrium short-selling is concentrated on asset 1, there are more sellers of this asset than in the symmetric equilibrium. There are also more buyers because of the short-sellers who need to buy the asset back. Conversely, asset 2 attracts fewer buyers and sellers than in the symmetric equilibrium.

The lending fee of asset 1 is higher than in the symmetric equilibrium because of two effects. First, because there are more buyers and sellers of asset 1, a short-sale is easier to execute, and

[29] Equations (25) and (26) ensure that arbitrage portfolios are suboptimal for arbitrageurs, i.e., average-valuation agents with *no* initial position. They do not apply, however, to average-valuation agents with "inherited" positions. Consider, for example, a low-valuation agent with a short position in asset 1, who reverts to average valuation. The agent can unwind the short position by trading with a seller of asset 1, but might also accept to trade with a seller of asset 2. This would hedge the short position, lowering the cost of waiting for a seller of asset 1.

In our analysis, we rule out such strategies by assuming that arbitrage portfolios can be held only by arbitrageurs. This is partly for simplicity, to keep agents' life-cycles manageable. One could also argue that many investors do not engage in such strategies because of costs to managing multiple positions, settlement costs, etc. (These costs could be smaller for sophisticated arbitrageurs.) Needless to say, it would be desirable to relax this assumption.

the short-selling surplus is higher. Moreover, lenders of asset 1 are in better position to bargain for this surplus because they do not have to compete with lenders of asset 2.

To explain the price results, we recall that prices differ from the Walrasian benchmark because of a liquidity discount, a bargaining discount, and a specialness premium. In the asymmetric equilibrium, asset 1's liquidity discount is smaller than in the symmetric equilibrium because there are more buyers. Moreover, asset 1's specialness premium is higher because of the higher lending fee. Conversely, asset 2's liquidity discount is higher than in the symmetric equilibrium, and its specialness premium is zero. Therefore, absent the bargaining discount, i.e., when the buyers' bargaining power ϕ is zero, asset 1 trades at a higher price and asset 2 at a lower price relative to the symmetric equilibrium.

Perhaps the most surprising result of Proposition 6 is that both assets can trade at a higher price relative to the symmetric equilibrium. Thus, the bargaining discount can reverse the effects of liquidity and specialness. To explain the intuition, we recall that short-sellers exit the seller pool faster when the asset they have borrowed has a larger buyer pool. This occurs in the asymmetric equilibrium because asset 1 has more buyers than either asset in the symmetric equilibrium. Therefore, there are fewer short-sellers in the asymmetric equilibrium, and the aggregate seller pool can be smaller. This can, in turn, worsen the buyers' bargaining position and raise the prices of both assets.

To measure social welfare, we add the utilities V_τ of all agents, discounting those of future entrants at the interest rate r. From the Bellman equations of Section 4.4, an agent's utility is equal to the PV of the flow benefits derived over the agent's lifetime. Therefore, social welfare is equal to the PV of the flow benefits derived by all agents. In the proof of Proposition 6 we show that welfare depends on the equilibrium allocation through

$$\sum_{i=1}^{2} \left[\mu_{\underline{n}i}(\overline{x} + \underline{x} - 2y) - \mu_{\overline{s}i}\overline{x} - \mu_{\underline{b}i}(2y - \overline{x}) \right]. \tag{27}$$

The first term inside the summation corresponds to the gains from trade between high- and low-valuation agents, achieved through short-sales. The extent of short-sales is given by the measure $\mu_{\underline{n}i}$ of low-valuation non-searchers. The last two terms correspond to inefficiencies arising because some average-valuation agents hold positions that are no longer optimal. These agents are either sellers $\overline{s}i$ seeking to unwind a long position, or buyers $\underline{b}i$ seeking to unwind a short position.

When search frictions are small, the measure $\mu_{\overline{s}i}$ converges to zero, while $\sum_{i=1}^{2} \mu_{\underline{n}i}$ converges to the measure $\underline{F}/\underline{\kappa}$ of low-valuation agents. Therefore, welfare depends on the equilibrium allocation

only through the measure $\sum_{i=1}^{2} \mu_{\underline{b}i}$ of buyers seeking to unwind short positions. In the asymmetric equilibrium these buyers can trade faster because asset 1 has more sellers than either asset in the symmetric equilibrium. Therefore, $\sum_{i=1}^{2} \mu_{\underline{b}i}$ is lower and social welfare higher.

5 Extensions

5.1 Different Supplies

In this section we consider the case where asset supplies differ. Without loss of generality, we take asset 1 to be in larger supply, i.e., $S_1 > S_2$.

Proposition 7 *Assume Equation (4). As λ and ν become large:*

(i) An equilibrium where low-valuation agents short-sell both assets exists for a set of values of $S_1 - S_2$ that converges to $\{0\}$.

(ii) An equilibrium where short-selling is concentrated on asset 1 exists for all values of $S_1 - S_2$.

(iii) An equilibrium where short-selling is concentrated on asset 2 exists for a set of values of $S_1 - S_2$ that converges to $[0, \hat{S}]$ with $\hat{S} > 0$.

(iv) Social welfare is higher when short-selling is concentrated on asset 1 rather than asset 2.

Proposition 7 shows that asset supply is a powerful device in selecting among the equilibria of Section 4. For small search frictions, the symmetric equilibrium becomes knife-edge, existing only when asset supplies are very close. The intuition is the asset in larger supply (asset 1) has a larger seller pool because it has a larger pool of lenders who can revert to average valuation. This makes it more attractive to short-sellers because they can unwind a position more easily. When search frictions are small, short-sellers can afford to wait for asset 1 in the repo market, declining to borrow asset 2, and this eliminates the symmetric equilibrium.

The asymmetric equilibria are not knife-edge. In particular, short-selling can be concentrated on the smaller-supply asset 2 even when supplies are not very close. Intuitively, while asset 2 has a smaller pool of lenders who can revert to average valuation, it can have a larger overall seller pool because of the short-sellers. This makes it more attractive to short-sell and reinforces the

equilibrium. Of course, short-sellers can compensate for the difference in supplies only when this difference is not too large. Otherwise, short-selling can only be concentrated on asset 1. Asset 1's seller pool in this equilibrium is larger than asset 2's in the equilibrium where asset 2 attracts the short-sellers. Thus, when short-selling is concentrated on asset 1, short-sellers can unwind their positions more easily and social welfare is higher.

Proposition 7 can reconcile our theory based on multiple equilibria with the empirical fact that liquidity in the US Treasury market concentrates in the just-issued bond. Indeed, a commonly-held view is that a bond's effective supply decreases over time as the bond becomes "locked away" in the portfolios of long-horizon investors (see Amihud and Mendelson [1991]). Our theory does not capture this effect because it focuses on steady-states and assumes equal horizons for all high-valuation investors. It suggests, however, that because off-the-run bonds are in smaller effective supply, they are less likely to attract short-sellers and for that reason less liquid.

5.2 Market Integration

We next relax the constraint that short-sellers can deliver only the asset they borrowed. In the Treasury market this could be achieved if on- and off-the-run bonds are standardized in terms of their maturity dates. For example, a two-year bond could be designed to mature on exactly the same date as a previously-issued five-year bond. The two bonds could then be made "fungible," assigned the same CUSIP number, and be identical for delivery purposes. Bennett et al. [2000] propose specific measures to implement this outcome, arguing that it would enhance the liquidity of the Treasury market.

When short-sellers can deliver any asset, markets are effectively integrated as if there is a single asset in supply $2S$. In Proposition 8 we compare this outcome to the equilibria of Propositions 2 and 4.

Proposition 8 *Suppose that there is a single asset in supply* $2S$*. Then*

(i) The asset price is higher than in the symmetric equilibrium.

(ii) The asset price is higher than the price of asset 2 in the asymmetric equilibrium. It can be higher or lower than the price of asset 1 and the average price of the two assets.

(iii) Social welfare is higher than in the symmetric and asymmetric equilibria.

Under market integration, each asset has more buyers than in the symmetric and asymmetric equilibria. Therefore, the liquidity and bargaining discounts are smaller. The specialness premium tends to be larger because market integration increases the short-selling surplus (by facilitating delivery), thus increasing the fee that lenders can extract. Offsetting this effect, is that lenders of asset 1 in the asymmetric equilibrium are in better position to bargain for the surplus because they do not have to compete with lenders of asset 2. This can generate the surprising result that the price under market integration can be lower than the average price in the asymmetric equilibrium. Social welfare is always higher, however, because short-sellers can deliver more easily.

An interesting implication of Proposition 8 is that while social welfare is always maximized under market integration, government revenue can be maximized in the asymmetric equilibrium. This suggests that in some circumstances, a revenue-maximizing Treasury might have no incentive to relax the constraint that short-sellers can only deliver the asset they borrowed.

6 Calibration

In this section we perform a calibration exercise, and show that our model can generate significant price effects even for short search times. For the calibration we extend the model to more than two assets. This provides a more accurate description of the US Treasury market, where there is one on-the-run and multiple off-the-run securities for each maturity range. With multiple assets there is again an equilibrium where short-sellers concentrate in one asset, e.g., asset 1. To compute this equilibrium for the purpose of calibration, we do not rely on the asymptotic closed-form solutions of Section 4. Instead, we use a simple numerical algorithm that solves the exact system of equations and checks that arbitrage is unprofitable. Table 1 lists our chosen values for the exogenous parameters, and Tables 2 and 3 list the calibration results.

We set the number of assets to $I = 20$, consistent with the fact that on-the-run bonds account for about 5% of the Treasury market capitalization (Dupont and Sack [1999]). We assume that all assets are in identical supply S. We normalize the total supply IS to one, without loss of generality: Equations (7)-(11) and (B.3)-(B.7) show that if $(S, \overline{F}, \underline{F}, 1/\lambda, 1/\nu)$ are scaled by the same factor, the meeting intensities of each investor type stay the same, and therefore prices and utilities do not change.

As in the case of two assets, we assume that demand exceeds supply, i.e., $\overline{F}/\overline{\kappa} > IS + \underline{F}/\underline{\kappa}$. We select $(\overline{F}, \underline{F})$ to make this an approximate equality; otherwise for small frictions, search times

for sellers would be much shorter than for buyers. We use the second degree of freedom in $(\overline{F}, \underline{F})$ to match the level of short-selling activity. Namely, in our calibration the amount of ongoing repo agreements for asset 1 is about seven times the asset's issue size (Table 2), which is within reasonable range.[30]

The expected investment horizons $1/\overline{\kappa}$ and $1/\underline{\kappa}$ are chosen to match turnover. Sundaresan [2002] and Strebulaev [2002] report that on-the-run bonds trade about ten times more than their off-the-run counterparts. Since the entire stock of Treasury securities turns over in less than three weeks (Dupont and Sack [1999]), on-the-run bonds turn over in about two-thirds of a day, and off-the-run bonds in about 125 days.[31] In our model the turnover of off-the-run bonds is generated by high-valuation investors. We let $1/\overline{\kappa} = 0.5$ years, i.e., 125 trading days, implying a turnover time of about the same (Table 2). The turnover of on-the-run bonds is generated mainly by short-sellers. We let $1/\underline{\kappa} = 0.025$ years, i.e., about six trading days. Such a short horizon could be reasonable for dealers in corporate bonds or mortgage-backed securities who have transitory needs to hedge inventory. For our chosen value of $\underline{\kappa}$, asset 1 turns over in 0.88 days, and its volume relative to the aggregate of the other assets is 7.5 (Table 2).[32] This is lower than the actual value of ten, but one could argue that short-selling is not the only factor driving the large relative volume of on-the-run bonds. Furthermore, raising the relative volume by increasing $\underline{\kappa}$ would strengthen our results because the lending fee would increase.

The parameters λ and ν are chosen to generate short search times, as reported in Table 2. Assuming ten trading hours per day,[33] most search times are in the order of a few hours or less, significantly smaller than the standard settlement time. It takes 12 minutes to sell the "on-the-run" asset 1 and 2.7 hours to buy it. Each "off-the-run" asset $i \in \{2, .., I\}$ can be sold in 2.8 hours and bought in 2.2 days. The time to buy might seem long, but is not unreasonable given that off-the-run bonds are often viewed as hard to locate. Furthermore, in our model all off-the-run assets are perfect substitutes for their buyers, who are the high-valuation agents. Therefore, a buyer's effective search time does not exceed $2.2/(I-1) = 0.11$ days. Finally, it takes 42 minutes to borrow

[30] For example, on February 2, 2005, primary dealers reported asset loans of about $2 trillion (New York Fed website, `www.ny.frb.org/markets/gsds/search.cfm`). Since the Treasury market is worth about $4 trillion, of which 5% are on-the-run bonds, the amount of repo agreements exceeds the market value of on-the-run bonds by about $2/(4 \times 5\%) = 10$. We select a number below ten to account for repo activity in off-the-run bonds. A higher number would strengthen our results because the lending fee would increase.

[31] Suppose, for example, that the average Treasury security turns over in twelve trading days. Since on-the-run bonds account for about 5% of market capitalization and 10/11 of trading volume, they turn over in $5\% \times 12/(10/11) = 0.66$ days, while off-the-run bonds turn over in $95\% \times 12/(1/11) = 125.4$ days.

[32] The six-day expected horizon of short-sellers is approximately equal to the turnover time of the asset supply that they generate $(\underline{F}/\underline{\kappa})$. This supply is about seven times the issue size S, and turns over seven times more slowly.

[33] US Treasury securities are traded round the clock in New York, London, and Tokyo. However, Fleming [1997] reports that 94% of the trading takes place in New York from 7:30am to 5:30pm.

Table 1: Parameter Values used in the Numerical Example.

Parameters		Value
Number of assets	I	20
Supply of each asset	S	0.05
Flow of high-valuation investors	$\overline{F}$	2.7
Flow of low-valuation investors	$\underline{F}$	13.6
Switching intensity of high-valuation investors	$\overline{\kappa}$	2
Switching intensity of low-valuation investors	$\underline{\kappa}$	40
Contact intensity in spot market	λ	10^6
Contact intensity in repo market	ν	7.5×10^4
Bargaining power of a buyer	ϕ	0.5
Bargaining power of a lender	θ	0.5
Riskless rate	r	4%
Dividend rate	δ	1
Hedging benefit of high-valuation investors	$\overline{x}$	0.4
Cost of risk bearing	y	0.5

asset 1 in the repo market and 8.7 hours to lend it. The time to lend the on-the-run asset might seem long but could be interpreted as an average across asset owners, some of whom do not engage in asset lending in practice.

The parameters ϕ and θ are set to 0.5 so that all agents are symmetric. The riskless rate r is set to 4%, consistent with Ibbotson [2004]'s average T-bill rate of 3.8% during the period 1926-2002. Given that prices and lending fees are linear in $(\delta, \overline{x}, \underline{x}, y)$, we let $\delta = 1$ and report relative prices (e.g., δ/p, w/p). The parameters $\overline{x}$ and y are selected based on assets' risk premia, measured by the difference $\delta/p_i - r$ between expected returns and the riskless rate. We assume that $\overline{x} < y$, so that the Walrasian price $(\delta + \overline{x} - y)/r$ incorporates a positive premium. We also assume that $y < \delta$, so that risk premia do not result in negative prices: the lowest possible price is $(\delta - y)/r$, the PV of the average-valuation agents' certainty equivalent. Our chosen values of $\overline{x}$ and y generate risk premia of about 2-2.5%, which are within reasonable range for government bonds. (For example, Ibbotson [2004] reports that long-term Treasuries returned 1.9% per year above bills during the period 1926-2002.[34]) To select $\underline{x}$, we note that Equation (5) suggests the restriction $\underline{x} \leq 4y - \overline{x} = 1.6$, because otherwise low-valuation agents could prefer to short more than one share. Moreover, our numerical calculations indicate that $\underline{x}$ must exceed 0.97, so that the lending fee is large enough to preclude arbitrage. We therefore assume $0.97 \leq \underline{x} \leq 1.6$, and report results for the two extreme values.

[34]Of course, this is only suggestive since in our model risk arises because of asset payoffs and not interest rates.

Table 2: Numerical Results: Search Times and Turnover.

Variable		Value
Average time to sell asset 1	$1/(\lambda\mu_{b1})$	0.02 days
Average time to buy asset 1	$1/(\lambda\mu_{s1})$	0.27 days
Average time to sell asset $i \in \{2,..,I\}$	$1/(\lambda\mu_{bi})$	0.28 days
Average time to buy asset $i \in \{2,..,I\}$	$1/(\lambda\mu_{si})$	2.20 days
Average time to borrow asset 1	$1/(\lambda\mu_{\overline{\ell}1})$	0.07 days
Average time to lend asset 1	$1/(\lambda\mu_{\underline{bo}})$	0.87 days
Time to turn over stock of asset 1	$S/(\lambda\mu_{b1}\mu_{s1})$	0.88 days
Time to turn over stock of asset $i \in \{2,..,I\}$	$S/(\lambda\mu_{bi}\mu_{si})$	125.28 days
Volume of asset 1 vs. aggregate of assets $i \in \{2,..,I\}$	$(\lambda\mu_{b1}\mu_{s1})/((I-1)\lambda\mu_{bi}\mu_{si})$	7.50
Repo agreements for asset 1 relative to issue size	$\mu_{\overline{n}1}/S$	7.03

Table 3: Numerical Results: Prices and Lending Fees.

Variable		$\underline{x} = 0.97$	$\underline{x} = 1.6$
Expected return of asset 1	δ/p_1	6.48%	6.02%
Expected return of asset $i \in \{2,..,I\}$	δ/p_i	6.52%	6.53%
Spread	$\delta/p_i - \delta/p_1$	4bp	51bp
Lending fee	w_1/p_1	3bp	35bp

Table 3 reports the prices and lending fees. When $\underline{x}$ is equal to its lowest value of 0.97, the effects are quite small: assets' expected returns differ by 4bp, and specialness is 3bp. When, however, $\underline{x}$ is equal to its highest value of 1.6, the effects are large and consistent with empirical findings. In particular, the 51bp difference in expected returns is consistent with Warga [1992], who reports that on-the-run portfolios return 55bp below matched off-the-run portfolios.[35] Moreover, the lending fee of 35bp is consistent with Duffie [1996], who reports a specialness difference of 40bp between on- and off-the-run bonds.[36]

[35]Some studies find smaller effects. For example, Goldreich et al. [2002] report that on-the-run bonds yield 1.5bp below off-the-run bonds, and Fleming [2003] reports 5.6bp. These papers, however, focus on bonds with a long time to maturity, for which the three-month convenience yield of being on-the-run has only a small effect on the yield to maturity. Warga [1992] compares the returns of on- and off-the-run bond portfolios rather than their yields to maturity. This isolates the on-the-run convenience yield in exactly the same way as in this paper. Amihud and Mendelson [1991] compare yields to maturity, but can isolate the convenience yield because they focus on securities with very short times to maturity. They find that Treasury bills maturing in less than six months yield 38bp below comparable Treasury notes.

[36]The expected return spread $\delta/p_i - \delta/p_1$ in Table 3 is greater than the lending fee w_1/p_1. This suggests an arbitrage strategy of shorting \$1 of asset 1, paying the lending fee, and buying \$1 of asset 2. The payoff of this strategy is risky, however, because the assets are held in different quantities. Adjusting for risk amounts to calculating the marginal utility flow $(\delta - y)/p_i - (\delta - y)/p_1 - w_1/p_1$ that an arbitrageur would derive, which turns out to be negative.

The large price effects are in spite of the short search times. The transaction costs implicit in these times are, in fact, very small. For example, the cost incurred by a high-valuation buyer is not to receive the hedging benefit $\overline{x}$ while searching. With a search time not exceeding 0.11 days, i.e., 0.11/250 of a year, the search cost is a fraction $\overline{x} \times (0.11/250) \times (1/6.53\%) = 1.1 \times 10^{-5}$ of the price, i.e., 0.11 cents per \$100 transaction value. Likewise, the search cost of a low-valuation agent seeking to borrow asset 1 in the repo market is not to receive the hedging benefit $\underline{x}$. When $\underline{x}$ is equal to its highest value of 1.6, the cost is a fraction $\underline{x} \times (0.07/250) \times (1/6.53\%) = 2.9 \times 10^{-5}$ of the price, i.e., 0.29 cents per \$100 transaction value. Such costs are smaller than the average bid-ask spread in the Treasury market, which is 1.1 cent (Dupont and Sack [1999]). While this raises the question of what drives the bid-ask spread, it also shows that the large price effects in our model are driven by very small transaction costs.[37]

Small transaction costs imply that most of the return spread is due to the specialness premium. Generalizing the decomposition in Section 4, we can show that when $\underline{x} = 1.6$ the specialness premium accounts for 99% of the spread while the liquidity premium for only 1%. Of course, this does not mean that liquidity does not matter; it rather means that liquidity can have large effects through specialness.

7 Empirical Implications

A basic implication of our theory is that liquidity and specialness are both generated by short-selling activity. Thus, on-the-run premia and specialness should be increasing in the extent of short-selling (represented by the short-seller flow $\underline{F}$), and be zero in markets without short-sellers. Furthermore, because short-sellers concentrate in the on-the-run bond, our theory implies that they should account for a larger fraction of on-the-run trading volume, relative to off-the-run.

Several empirical studies document a systematic relationship between on-the-run premia, specialness, and short-selling activity. Jordan and Jordan [1997] provides a case study where short-seller demand for a particular Treasury note generated a large price premium. Krishnamurthy [2002] measures short-seller demand by the issuance of corporate and agency bonds, arguing that dealers short Treasuries to hedge their inventories. He finds that issuance is positively related to the on-the-run premium. Graveline and McBrady [2004] emphasize the role of short-seller demand using a conceptual framework very similar to ours. They construct several measures of demand,

[37] Our model could generate larger transaction costs if agents had to incur a monetary cost of search (e.g., pay a broker), in addition to not receiving hedging benefits.

attempting to get both at the hedging and the speculative component. They find that short-seller demand is the strongest determinant of specialness once variation related to the auction cycle is taken out. Moulton [2004] also finds evidence linking short-selling demand to specialness.

Another determinant of on-the-run premia and specialness is the supply of lendable securities. Cornell and Shapiro [1989] and Jordan and Jordan [1997] provide case studies where large price premia were generated by a short-squeeze or a large investor's unwillingness to lend, respectively. Krishnamurthy [2002] finds that on-the-run premia are negatively related to issue size, and Graveline and McBrady [2004] and Moulton [2004] find a negative relationship between issue size and specialness. On the other hand, a commonly held view is that on-the-run bonds are on special more frequently than off-the-run bonds because of their larger effective supply. Our model can reconcile these observations because it implies that the effect of issue size S_1 on specialness is non-monotonic. A decrease in S_1 initially increases specialness through a scarcity effect: because the asset becomes harder to find in the repo market, lenders have more bargaining power, and the lending fee increases. There is, however, a countervailing liquidity effect: because the pool of sellers becomes smaller, short-sellers have greater difficulty buying back the asset, and this tends to reduce the short-selling surplus and the lending fee. If S_1 becomes sufficiently small relative to S_2, the short-selling surplus becomes negative, short-sellers concentrate in the other asset, and specialness drops to zero.

An additional determinant of on-the-run premia and specialness is the investors' demand for liquidity. Krishnamurthy [2002] measures liquidity demand by the spread between Commercial Paper and T-Bills, and shows that it is positively related to the on-the-run premium. Graveline and McBrady [2004] and Moulton [2004] show that several measures of liquidity demand are positively related to specialness. In our model, demand for liquid assets can be measured by the expected investment horizon $1/\overline{\kappa}$ of the agents holding long positions. A decrease in horizon, holding the total measure $\overline{F}/\overline{\kappa}$ of longs constant, tends to increase specialness because of two effects. First, because longs call their asset loans more frequently, short-sellers return more frequently to the repo market and bid up the lending fee. Second, because longs turn more frequently into sellers, the seller pool is larger. Therefore, short-sellers can buy back the asset more easily and have a larger surplus.

8 Conclusion

This paper proposes a search-based theory of the on-the-run phenomenon. We take the view that liquidity and specialness are not independent explanations of this phenomenon, but can be explained simultaneously by short-selling activity. Short-sellers in our model can endogenously concentrate in one of two identical assets, because of search externalities and the constraint that they must deliver the asset they borrowed. That asset enjoys both greater liquidity, measured by search times, and a higher lending fee ("specialness"). Moreover, liquidity and specialness translate into price premia which are consistent with no-arbitrage. We derive closed-form solutions in the realistic case of small frictions, and show that a calibration can generate effects of the observed magnitude. Our model can shed light on additional aspects of the on-the-run phenomenon, such as the effects of issue size and market integration, and the apparent puzzle that off-the-run bonds are cheap yet "scarce."

While our analysis is motivated from the government-bond market, some lessons can be more general. Perhaps the main lesson concerns the law of one price - a fundamental tenet of Finance. We show that this law can be violated in a significant manner in a model where all agents are rational but the trading mechanism is not Walrasian. Our search-based trading mechanism is of course an idealization, but it captures the bilateral nature of trading in over-the-counter markets. Furthermore, the search times that are needed to generate significant price differentials are small, in the order of a few hours. For such times, it is unclear whether the search framework is a worse description of over-the-counter markets than a Walrasian auction, which assumes multilateral trading.

A Walrasian Equilibrium

In this section we prove Proposition 1. An agent maximizes his intertemporal utility (2) subject to the budget constraint

$$dW_t = \left[rW_t - c_t + \sum_{i=1}^{2} (\delta - rp_i) z_{it} \right] dt + \left[\sigma \sum_{i=1}^{2} z_{it} + \rho_t \sigma_e \right] dB_t + \sigma_e \sqrt{1-\rho_t^2}\, dZ_t$$

and the transversality condition

$$\lim_{T \to \infty} E\left[\exp(-AW_T - \beta T) \right] = 0,$$

where W_t is the wealth, z_{it} the number of shares invested in asset $i \in \{1,2\}$, and $A \equiv ra$. The agent's controls are the consumption $c \in \mathbb{R}$ and the investments $z_1, z_2 \in \mathbb{Z}$. Obviously, if $p_1 \neq p_2$ the agent can achieve infinite utility by demanding an infinite amount of assets, contradicting equilibrium. Thus, in equilibrium p_1 and p_2 must be equal. Denoting their common value by p, and the aggregate investment in the risky assets by $z \equiv z_1 + z_2$, we can write the budget constraint as

$$dW_t = \left[rW_t - c_t + (\delta - rp) z_t \right] dt + \left[\sigma z_t + \rho_t \sigma_e \right] dB_t + \sigma_e \sqrt{1-\rho_t^2}\, dZ_t.$$

The agent's value function $J(W_t, \rho_t)$ satisfies the Hamilton-Jacobi-Bellman (HJB) equation

$$0 = \sup_{(c,z) \in \mathbb{R} \times \mathbb{Z}} \left\{ -\exp(-\alpha c) + \mathcal{D}^{(c,z)} J(W,\rho) - \beta J(W,\rho) \right\}, \tag{A.1}$$

where

$$\begin{aligned} \mathcal{D}^{(c,z)} J(W,\rho) \;\equiv\; & J_W(W,\rho) \left[rW - c + (\delta - rp) z \right] + \frac{1}{2} J_{WW}(W,\rho) \left[\sigma^2 z^2 + 2\rho\sigma\sigma_e z + \sigma_e^2 \right] \\ & + \kappa(\rho) \left[J(W,0) - J(W,\rho) \right], \end{aligned}$$

and where the transition intensity $\kappa(\rho)$ is equal to $\overline{\kappa}$ for $\rho = \overline{\rho}$, and $\underline{\kappa}$ for $\rho = \underline{\rho}$. We guess that $J(W,\rho)$ takes the form

$$J(W,\rho) = -\frac{1}{r} \exp\left[-A[W + V(\rho)] + \frac{r - \beta + \frac{A^2 \sigma_e^2}{2}}{r} \right],$$

for some function $V(\rho)$. Replacing into Equation (A.1), we find that the optimal consumption is

$$c(\rho) = r[W + V(\rho)] - \frac{r - \beta + \frac{A^2\sigma_\varepsilon^2}{2}}{A}$$

and the optimal investment satisfies

$$z(\rho) \in \text{argmax}_{z\in\mathbb{Z}}\{C(\rho, z) - rpz\} \equiv Z(\rho)$$

for the incremental certainty equivalent $C(\rho, z)$ introduced in Section 3. Plugging $c(\rho)$ back into Equation (A.1), we find that Equation (A.1) is satisfied iff

$$0 = -rV(\rho) + \max_{z\in\mathbb{Z}}\{C(\rho, z) - rpz\} + \kappa(\rho)\frac{1 - e^{-A(V(0)-V(\rho))}}{A}. \tag{A.2}$$

Equation (A.2) implies that $V(0) = \max_z\{C(\rho, z) - rpz\}/r$. Moreover, given $V(0)$, the equations for $V(\overline{\rho})$ and $V(\underline{\rho})$ are in only one unknown, and it is easy to check that they have a unique solution.

We next determine the equilibrium value of p. Because each type-ρ agent holds a position $z(\rho) \in Z(\rho)$, the average position $z_m(\rho)$ of these agents is in the convex hull of $Z(\rho)$. Market clearing requires that $z_m(0) = 0$ because average-valuation agents in infinite measure. It also requires that

$$\frac{\overline{F}}{\overline{\kappa}}z_m(\overline{\rho}) + \frac{\underline{F}}{\underline{\kappa}}z_m(\underline{\rho}) = \sum_{i=1}^{2} S_i. \tag{A.3}$$

Because the function $z \to C(\rho, z) - rpz$ is strictly concave, the set $Z(\rho)$ consists of either one or two elements. If there exists a z such that

$$C(\rho, z) - rpz > \max\left\{C(\rho, z+1) - rp(z+1), C(\rho, z-1) - rp(z-1)\right\}, \tag{A.4}$$

then this z is unique and $Z(\rho) = \{z\}$. Otherwise, there exists a unique z such that

$$C(\rho, z) - rpz = C(\rho, z+1) - rp(z+1), \tag{A.5}$$

and $Z(\rho) = \{z, z+1\}$. Using Equation (5) and the first-order conditions (A.4) and (A.5), it is easy to check that for $p = (d + \overline{x} - y)/r$, $Z(\overline{\rho}) = \{0, 1\}$, $Z(\underline{\rho}) = \{-1\}$, and $Z(0) = \{0\}$. Equation (A.3) follows then from (4), implying that $p = (\delta + \overline{x} - y)/r$ is an equilibrium price. It is the unique equilibrium price because if $p > (\delta + \overline{x} - y)/r$ then no agent would choose $z > 0$, and if $p < (\delta + \overline{x} - y)/r$ then high-valuation agents would choose $z \geq 1$, while other agents would choose at least as much as for $p = (\delta + \overline{x} - y)/r$. ∎

B Demographics

B.1 Inflow-Outflow Equations

The inflows and outflows for each agent type are as follows:

Lenders $\overline{\ell}i$: The inflow is the sum of $\lambda\mu_{\overline{b}}\mu_{si}$ because some high-valuation buyers meet sellers, and a flow f_i from the high-valuation non-searchers. The outflow is the sum of $\overline{\kappa}\mu_{\overline{\ell}i}$ because some lenders switch to average valuation and become sellers, and $\nu_i\mu_{\underline{bo}}\mu_{\overline{\ell}i}$ because some lenders meet borrowers and become high-valuation non-searchers. Thus,

$$\lambda\mu_{\overline{b}}\mu_{si} + f_i = \overline{\kappa}\mu_{\overline{\ell}i} + \nu_i\mu_{\underline{bo}}\mu_{\overline{\ell}i}. \tag{B.1}$$

High-valuation non-searchers $\overline{n}i$: The inflow is $\nu_i\mu_{\underline{bo}}\mu_{\overline{\ell}i}$ from the lenders. The outflow is the sum of f_i, and $\overline{\kappa}\mu_{\overline{n}i}$ because some high-valuation non-searchers revert to average valuation and either become sellers (flow $\overline{\kappa}\mu_{\underline{s}i}$) or seize the collateral and exit the market (flow $\overline{\kappa}(\mu_{\underline{n}i} + \mu_{\underline{b}i})$). Thus,

$$\nu_i\mu_{\underline{bo}}\mu_{\overline{\ell}i} = f_i + \overline{\kappa}\mu_{\overline{n}i}. \tag{B.2}$$

Sellers $\overline{s}i$: The inflow is the sum of $\overline{\kappa}\mu_{\overline{\ell}i}$ from the lenders, and $\overline{\kappa}\mu_{\underline{s}i}$ from the high-valuation non-searchers. The outflow is $\lambda\mu_{bi}\mu_{\overline{s}i}$ because some sellers meet buyers and exit the market. Thus,

$$\overline{\kappa}\mu_{\overline{\ell}i} + \overline{\kappa}\mu_{\underline{s}i} = \lambda\mu_{bi}\mu_{\overline{s}i}. \tag{B.3}$$

Borrowers $\underline{bo}$: The inflow is the sum of $\underline{F}$ because of the new entrants, and $\sum_{i=1}^{2}\overline{\kappa}(\mu_{\underline{s}i} + \mu_{\underline{n}i})$ because of the low-valuation sellers and non-searchers who are called to deliver the asset. The outflow is the sum of $\underline{\kappa}\mu_{\underline{bo}}$ because some borrowers revert to average valuation and exit the market, and $\sum_{i=1}^{2}\nu_i\mu_{\underline{bo}}\mu_{\overline{\ell}i}$ because some borrowers meet lenders and become low-valuation sellers. Thus,

$$\underline{F} + \sum_{i=1}^{2}\overline{\kappa}(\mu_{\underline{s}i} + \mu_{\underline{n}i}) = \underline{\kappa}\mu_{\underline{bo}} + \sum_{i=1}^{2}\nu_i\mu_{\underline{bo}}\mu_{\overline{\ell}i}. \tag{B.4}$$

Low-valuation sellers $\underline{s}i$: The inflow is $\nu_i\mu_{\underline{bo}}\mu_{\overline{\ell}i}$ from the borrowers. The outflow is the sum of $\overline{\kappa}\mu_{\underline{s}i}$ because some low-valuation sellers are called to deliver the asset and become borrowers, $\underline{\kappa}\mu_{\underline{s}i}$

because some low-valuation sellers revert to average valuation and exit the market, and $\lambda\mu_{bi}\mu_{\underline{s}i}$ because some low-valuation sellers meet buyers and become low-valuation non-searchers. Thus,

$$\nu_i\mu_{\underline{b}o}\mu_{\overline{\ell}i} = \overline{\kappa}\mu_{\underline{s}i} + \underline{\kappa}\mu_{\underline{s}i} + \lambda\mu_{bi}\mu_{\underline{s}i}. \tag{B.5}$$

Low-valuation non-searchers $\underline{n}i$: The inflow is $\lambda\mu_{bi}\mu_{\underline{s}i}$ from the low-valuation sellers. The outflow is the sum of $\overline{\kappa}\mu_{\underline{n}i}$ because some low-valuation non-searchers are called to deliver the asset and become borrowers, and $\underline{\kappa}\mu_{\underline{n}i}$ because some low-valuation non-searchers revert to average valuation and become buyers. Thus,

$$\lambda\mu_{bi}\mu_{\underline{s}i} = \overline{\kappa}\mu_{\underline{n}i} + \underline{\kappa}\mu_{\underline{n}i}. \tag{B.6}$$

Buyers $\underline{b}i$: The inflow is $\underline{\kappa}\mu_{\underline{n}i}$ from the high-valuation non-searchers. The outflow is the sum of $\overline{\kappa}\mu_{\underline{b}i}$ because some buyers are called to deliver the asset and exit the market, and $\lambda\mu_{\underline{b}i}\mu_{si}$ because some buyers meet sellers and exit the market. Thus,

$$\underline{\kappa}\mu_{\underline{n}i} = \overline{\kappa}\mu_{\underline{b}i} + \lambda\mu_{\underline{b}i}\mu_{si}. \tag{B.7}$$

We consider the system formed by the accounting equations (7) and (8), the market-clearing equations (9) and (10), and the inflow-outflow equations (11) and (B.3)-(B.7). The total number of equations is 18 (because some are for each asset), and the 18 unknowns are the measures of the 14 agent types and $\{\mu_{bi}, \mu_{si}\}_{i\in\{1,2\}}$. A solution to the system satisfies Equations (B.1) and (B.2), which is why we do not include them into the system. Indeed, adding Equations (B.5)-(B.7), and using Equation (10), we find

$$\nu_i\mu_{\underline{b}o}\mu_{\overline{\ell}i} = \overline{\kappa}\mu_{\underline{s}i} + \underline{\kappa}\mu_{\overline{n}i} + \lambda\mu_{\underline{b}i}\mu_{si}.$$

Therefore, Equation (B.2) holds with

$$f_i = \underline{\kappa}\mu_{\underline{s}i} + \lambda\mu_{\underline{b}i}\mu_{si}.$$

For this value of f_i, Equation (B.1) becomes

$$\lambda\mu_{bi}\mu_{si} + \underline{\kappa}\mu_{\underline{s}i} = \overline{\kappa}\mu_{\overline{\ell}i} + \nu_i\mu_{\underline{b}o}\mu_{\overline{\ell}i},$$

and is redundant because it can be derived by adding Equations (B.3) and (B.5).

To solve the system, we reduce it to a simpler one in the six unknowns $\mu_{\underline{bo}}$, $\mu_{\overline{b}}$, and $\{\mu_{bi}, \mu_{si}\}_{i\in\{1,2\}}$. Adding Equations (B.5) and (B.6), we find

$$\mu_{\underline{s}i} + \mu_{\underline{n}i} = \frac{\nu_i \mu_{\underline{bo}} \mu_{\overline{\ell}i}}{\overline{\kappa} + \underline{\kappa}}. \tag{B.8}$$

Plugging into equation (B.4), and using Equation (9), we find

$$\underline{F} = \underline{\kappa}\mu_{\underline{bo}} + \frac{\underline{\kappa}}{\overline{\kappa} + \underline{\kappa}} \sum_{i=1}^{2} \nu_i \mu_{\underline{bo}} (S - \mu_{si}). \tag{B.9}$$

Equations (B.5) and (9) imply that

$$\mu_{\underline{s}i} = \frac{\nu_i \mu_{\underline{bo}} (S - \mu_{si})}{\overline{\kappa} + \underline{\kappa} + \lambda\mu_{bi}}. \tag{B.10}$$

Equation (B.6) implies that

$$\mu_{\underline{n}i} = \frac{\lambda \mu_{\underline{s}i} \mu_{bi}}{\overline{\kappa} + \underline{\kappa}} \tag{B.11}$$

and Equation (B.7) implies that

$$\mu_{\underline{b}i} = \frac{\underline{\kappa}\mu_{\underline{n}i}}{\overline{\kappa} + \lambda\mu_{si}}. \tag{B.12}$$

Combining these equations to compute $\mu_{\underline{b}i}$, and using Equation (7), we find

$$\mu_{bi} = \mu_{\overline{b}} + \frac{\underline{\kappa}\lambda\mu_{bi}\nu_i\mu_{\underline{bo}}(S - \mu_{si})}{(\overline{\kappa} + \underline{\kappa})(\overline{\kappa} + \underline{\kappa} + \lambda\mu_{bi})(\overline{\kappa} + \lambda\mu_{si})}. \tag{B.13}$$

Noting that $\mu_{\overline{\ell}i} + \mu_{\underline{s}i} = S - \mu_{\overline{s}i}$, we can use Equation (B.3) to compute $\mu_{\overline{s}i}$:

$$\mu_{\overline{s}i} = \frac{\overline{\kappa}S}{\overline{\kappa} + \lambda\mu_{bi}}. \tag{B.14}$$

Adding Equations (B.10) and (B.14), and using Equation (8), we find

$$\mu_{si} = \frac{\overline{\kappa}S}{\overline{\kappa} + \lambda\mu_{bi}} + \frac{\nu_i \mu_{\underline{bo}} (S - \mu_{si})}{\underline{\kappa} + \overline{\kappa} + \lambda\mu_{bi}}. \tag{B.15}$$

The new system consists of Equations (11), (B.9), (B.13), and (B.15). These are six equations (because some are for each asset), and the six unknowns are $\mu_{\underline{bo}}$, $\mu_{\overline{b}}$, and $\{\mu_{bi}, \mu_{si}\}_{i\in\{1,2\}}$. Once this system is solved, the other measures can be computed as follows: $\mu_{\underline{s}i}$ from (B.10), $\mu_{\underline{n}i}$ from (B.11), $\mu_{\underline{b}i}$ from (B.12), $\mu_{\overline{s}i}$ from (B.14), $\mu_{\overline{\ell}i}$ from (9), and $\mu_{\overline{n}i}$ from (10).

To cover the case where search frictions are small, we make the change of variables $\varepsilon \equiv 1/\lambda$, $n \equiv \nu/\lambda$, $\alpha_i \equiv \nu_i/\nu$, $\gamma_{si} \equiv \lambda\mu_{si}$, and $\gamma_{\underline{bo}} \equiv \nu\mu_{\underline{bo}}$. Under the new variables, Equations (11), (B.9), (B.13), and (B.15) become

$$\overline{F} = \overline{\kappa}\mu_{\overline{b}} + \sum_{i=1}^{2} \mu_{\overline{b}}\gamma_{si}, \tag{B.16}$$

$$\underline{F} = \frac{\varepsilon\underline{\kappa}\gamma_{\underline{bo}}}{n} + \frac{\underline{\kappa}}{\overline{\kappa}+\underline{\kappa}} \sum_{i=1}^{2} \alpha_i\gamma_{\underline{bo}}(S - \varepsilon\gamma_{si}), \tag{B.17}$$

$$\mu_{bi} = \mu_{\overline{b}} + \frac{\underline{\kappa}\mu_{bi}\alpha_i\gamma_{\underline{bo}}(S-\varepsilon\gamma_{si})}{(\overline{\kappa}+\underline{\kappa})\left[\varepsilon(\overline{\kappa}+\underline{\kappa})+\mu_{bi}\right](\overline{\kappa}+\gamma_{si})}, \tag{B.18}$$

$$\gamma_{si} = \frac{\overline{\kappa}S}{\varepsilon\overline{\kappa}+\mu_{bi}} + \frac{\alpha_i\gamma_{\underline{bo}}(S-\varepsilon\gamma_{si})}{\varepsilon(\underline{\kappa}+\overline{\kappa})+\mu_{bi}}, \tag{B.19}$$

respectively.

B.2 Existence and Uniqueness

We next show that the system of Equations (B.16)-(B.19) has a unique symmetric solution when $\alpha_1 = \alpha_2 = 1$ (the "symmetric" case), and a unique solution when $\alpha_1 = 1$ and $\alpha_2 = 0$ (the "asymmetric" case). Using Equation (B.18) to eliminate $\gamma_{\underline{bo}}$ in Equation (B.19), we find

$$\gamma_{si} = \frac{\overline{\kappa}S}{\varepsilon\overline{\kappa}+\mu_{bi}} + (\mu_{bi} - \mu_{\overline{b}})\frac{(\overline{\kappa}+\underline{\kappa})(\overline{\kappa}+\gamma_{si})}{\underline{\kappa}\mu_{bi}}.$$

Multiplying by μ_{bi}, and setting $i = 1$, we find

$$\gamma_{s1}\mu_{\overline{b}} = \frac{\overline{\kappa}S\mu_{b1}}{\varepsilon\overline{\kappa}+\mu_{b1}} + (\mu_{b1} - \mu_{\overline{b}})\frac{\overline{\kappa}}{\underline{\kappa}}(\overline{\kappa}+\underline{\kappa}+\gamma_{s1}). \tag{B.20}$$

In the rest of the proof, we use Equations (B.16), (B.17), (B.18) for $i \in \{1,2\}$, and (B.19) for $i = 2$, to determine $\mu_{\overline{b}}$ and μ_{b1} as functions of $\gamma_{s1} \in (0, S/\varepsilon)$. We then plug these functions into Equation (B.20), and show that the resulting equation in the single unknown γ_{s1} has a unique solution.

We first solve for $\mu_{\overline{b}}$. In the asymmetric case, Equation (B.18) implies that $\mu_{b2} = \mu_{\overline{b}}$, Equation (B.19) implies that $\gamma_{s2} = \overline{\kappa}S/(\varepsilon\overline{\kappa} + \mu_{\overline{b}})$, and Equation (B.16) implies that

$$\overline{F} = \overline{\kappa}\mu_{\overline{b}} + \mu_{\overline{b}}\left(\gamma_{s1} + \frac{\overline{\kappa}S}{\varepsilon\overline{\kappa} + \mu_{\overline{b}}}\right). \tag{B.21}$$

The RHS of Equation (B.21) is (strictly) increasing in $\mu_{\overline{b}} \in (0, \infty)$, is equal to zero for $\mu_{\overline{b}} = 0$, and goes to ∞ for $\mu_{\overline{b}} \to \infty$. Therefore, Equation (B.21) has a unique solution $\mu_{\overline{b}} \in (0, \infty)$. This solution is decreasing in γ_{s1} because the RHS is increasing in γ_{s1}. In the symmetric case, Equation (B.16) implies that $\mu_{\overline{b}} = \overline{F}/(\overline{\kappa} + 2\gamma_{s1})$. This solution is again decreasing in γ_{s1}.

We next solve for μ_{b1}. Equation (B.17) implies that

$$\gamma_{\underline{bo}} = \frac{\underline{F}}{\frac{\varepsilon\underline{\kappa}}{n} + \frac{\underline{\kappa}}{\overline{\kappa}+\underline{\kappa}}\sum_{i=1}^{2}\alpha_i(S - \varepsilon\gamma_{si})} = \frac{\underline{F}}{\frac{\varepsilon\underline{\kappa}}{n} + \frac{\underline{\kappa}}{\overline{\kappa}+\underline{\kappa}}(1+\alpha_2)(S - \varepsilon\gamma_{s1})},$$

where the second step follows because in the symmetric case $\gamma_{s2} = \gamma_{s1}$ and in the asymmetric case $\alpha_2 = 0$. Plugging into Equation (B.18), setting $i = 1$, and dividing by μ_{b1}, we find

$$1 = \frac{\mu_{\overline{b}}}{\mu_{b1}} + \frac{(S - \varepsilon\gamma_{s1})n\underline{F}}{\left[\varepsilon(\overline{\kappa} + \underline{\kappa}) + \mu_{b1}\right](\overline{\kappa} + \gamma_{s1})\left[\varepsilon(\overline{\kappa} + \underline{\kappa}) + n(1+\alpha_2)(S - \varepsilon\gamma_{s1})\right]}. \tag{B.22}$$

The RHS of Equation (B.22) is decreasing in $\mu_{b1} \in (0, \infty)$, goes to ∞ for $\mu_{b1} \to 0$, and goes to zero for $\mu_{b1} \to \infty$. Therefore, Equation (B.21) has a unique solution $\mu_{b1} \in (0, \infty)$. This solution is decreasing in γ_{s1} because the RHS is decreasing in γ_{s1} and increasing in $\mu_{\overline{b}}$ (which is decreasing in γ_{s1}).

We next substitute $\mu_{\overline{b}}$ and μ_{b1} into Equation (B.20), and treat it as an equation in the single unknown γ_{s1}. To show uniqueness, we will show that the LHS is increasing in γ_{s1} and the RHS is decreasing. In the symmetric case, the LHS is equal to

$$\gamma_{s1}\mu_{\overline{b}} = \frac{\gamma_{s1}\overline{F}}{\overline{\kappa} + 2\gamma_{s1}},$$

and is increasing. In the asymmetric case, Equation (B.21) implies that the LHS is equal to

$$\gamma_{s1}\mu_{\overline{b}} = \overline{F} - \overline{\kappa}\mu_{\overline{b}} - \frac{\overline{\kappa}S\mu_{\overline{b}}}{\varepsilon\overline{\kappa} + \mu_{\overline{b}}},$$

and is increasing because $\mu_{\overline{b}}$ is decreasing in γ_{s1}. The first term in the RHS is increasing in μ_{b1}, and thus decreasing in γ_{s1}. To show that the second term is also decreasing, we multiply Equation (B.22) by $\mu_{b1}(\overline{\kappa}+\underline{\kappa}+\gamma_{s1})$:

$$(\mu_{b1}-\mu_{\overline{b}})(\overline{\kappa}+\underline{\kappa}+\gamma_{s1}) = \frac{\mu_{b1}(\overline{\kappa}+\underline{\kappa}+\gamma_{s1})(S-\varepsilon\gamma_{s1})n\underline{F}}{[\varepsilon(\overline{\kappa}+\underline{\kappa})+\mu_{b1}]\,(\overline{\kappa}+\gamma_{s1})\,[\varepsilon(\overline{\kappa}+\underline{\kappa})+n(1+\alpha_2)(S-\varepsilon\gamma_{s1})]}.$$

The RHS of this equation is decreasing in γ_{s1} because it is decreasing in γ_{s1} and increasing in μ_{b1} (which is decreasing in γ_{s1}). Therefore, the second term in the RHS of Equation (B.20) is decreasing in γ_{s1}.

To show existence, we note that for $\gamma_{s1}=0$, the LHS of Equation (B.20) is equal to zero, while the RHS is positive. Moreover, for $\gamma_{s1}=S/\varepsilon$, the LHS is equal to $S\mu_{\overline{b}}/\varepsilon$, while the RHS is equal to

$$\frac{\overline{\kappa}S\mu_{\overline{b}}}{\varepsilon\overline{\kappa}+\mu_{\overline{b}}} < \frac{S\mu_{\overline{b}}}{\varepsilon}$$

because $\mu_{b1}=\mu_{\overline{b}}$. Therefore, there exists a solution $\gamma_{s1}\in(0,S/\varepsilon)$.

B.3 Small Search Frictions

The case of small search frictions corresponds to small ε. Thus, the solution in this case is close to that for $\varepsilon=0$ provided that continuity holds. Our proof so far covers the case $\varepsilon=0$, except for existence. We next show that Equation (16) ensures existence for $\varepsilon=0$. We also compute the solution in closed form and show continuity.

To emphasize that $\varepsilon=0$ is a limit case, we use m and g instead of μ and γ. Equations (B.16)-(B.19) become

$$\overline{F} = \overline{\kappa}m_{\overline{b}}+\sum_{i=1}^{2} m_{\overline{b}}g_{si}, \tag{B.23}$$

$$\underline{F} = \frac{\underline{\kappa}}{\overline{\kappa}+\underline{\kappa}}\sum_{i=1}^{2}\alpha_i g_{\underline{bo}}S, \tag{B.24}$$

$$m_{bi} = m_{\overline{b}}+\frac{\underline{\kappa}\alpha_i g_{\underline{bo}}S}{(\overline{\kappa}+\underline{\kappa})(\overline{\kappa}+g_{si})}, \tag{B.25}$$

$$g_{si} = \frac{\overline{\kappa}S}{m_{bi}}+\frac{\alpha_i g_{\underline{bo}}S}{m_{bi}}. \tag{B.26}$$

We first solve the system of (B.23)-(B.26) in the symmetric case ($\alpha_1 = \alpha_2 = 1$), suppressing the asset subscript because of symmetry. Equation (B.24) implies that

$$g_{\underline{bo}} = \frac{(\overline{\kappa} + \underline{\kappa})\underline{F}}{2\underline{\kappa}S}, \tag{B.27}$$

Equation (B.26) implies that

$$g_s = \frac{\overline{\kappa}S + \frac{\overline{\kappa}+\underline{\kappa}}{2\underline{\kappa}}\underline{F}}{m_b}, \tag{B.28}$$

and Equation (B.23) implies that

$$m_{\overline{b}} = \frac{\overline{F}}{\overline{\kappa} + 2g_s}. \tag{B.29}$$

Substituting $g_{\underline{bo}}$, g_s, and $m_{\overline{b}}$ from Equations (B.27)-(B.29) into Equation (B.25), we find that m_b solves the equation

$$1 = \frac{\overline{F}}{\overline{\kappa}m_b + 2\overline{\kappa}S + \frac{\overline{\kappa}+\underline{\kappa}}{\underline{\kappa}}\underline{F}} + \frac{\underline{F}}{2\overline{\kappa}m_b + 2\overline{\kappa}S + \frac{\overline{\kappa}+\underline{\kappa}}{\underline{\kappa}}\underline{F}}. \tag{B.30}$$

This equation has a positive solution because of Equation (16).

We next consider the asymmetric case ($\alpha_1 = 1$, $\alpha_2 = 0$), and use $\hat{m}$ and $\hat{g}$ instead of m and g. Equation (B.25) implies that $\hat{m}_{b2} = \hat{m}_{\overline{b}}$, Equation (B.26) implies that

$$\hat{g}_{s2} = \frac{\overline{\kappa}S}{\hat{m}_{\overline{b}}}, \tag{B.31}$$

Equation (B.24) implies that

$$\hat{g}_{\underline{bo}} = \frac{(\overline{\kappa} + \underline{\kappa})\underline{F}}{\underline{\kappa}S}, \tag{B.32}$$

Equation (B.26) implies that

$$\hat{g}_{s1} = \frac{\overline{\kappa}S + \frac{\overline{\kappa}+\underline{\kappa}}{\underline{\kappa}}\underline{F}}{\hat{m}_{b1}}, \tag{B.33}$$

and Equation (B.23) implies that

$$\hat{m}_{\overline{b}} = \frac{\overline{F} - \overline{\kappa}S}{\overline{\kappa} + \hat{g}_{s1}}. \tag{B.34}$$

Substituting $\hat{g}_{\underline{b}o}$, $\hat{g}_{s1}$, and $\hat{m}_{\overline{b}}$ from Equations (B.32)-(B.34) into Equation (B.25), we find

$$\hat{m}_{b1} = \frac{\overline{F}}{\overline{\kappa}} - 2S - \frac{\underline{F}}{\underline{\kappa}}, \tag{B.35}$$

which is positive because of Equation (16).

To show continuity at $\varepsilon = 0$, we write Equation (B.20) as

$$\gamma_{s1}\mu_{\overline{b}} - \frac{\overline{\kappa}S\mu_{b1}}{\varepsilon\overline{\kappa} + \mu_{b1}} - (\mu_{b1} - \mu_{\overline{b}})\frac{\overline{\kappa}}{\underline{\kappa}}(\overline{\kappa} + \underline{\kappa} + \gamma_{s1}) = 0,$$

and denote by $R(\gamma_{s1}, \varepsilon)$ the RHS (treating $\mu_{\overline{b}}$ and μ_{b1} as functions of $(\gamma_{s1}, \varepsilon)$). Because $\mu_{\overline{b}}, \mu_{b1} > 0$ for $(\gamma_{s1}, \varepsilon) = (g_{s1}, 0)$ (symmetric case) and $(\gamma_{s1}, \varepsilon) = (\hat{g}_{s1}, 0)$ (asymmetric case), the functions $\mu_{\overline{b}}$ and μ_{b1} are continuously differentiable around that point, and so is the function $R(\gamma_{s1}, \varepsilon)$. Moreover, our uniqueness proof shows that the derivative of $R(\gamma_{s1}, \varepsilon)$ w.r.t. γ_{s1} is positive. Therefore, the Implicit Function Theorem ensures that for small ε, Equation (B.20) has a continuous solution $\gamma_{s1}(\varepsilon)$. Because of uniqueness, this solution coincides with the one that we have identified.

C Optimization

This appendix studies the stochastic control problem faced by an individual investor with CARA utility, in the search equilibrium of Section 4. We define the investor's problem, provide Hamilton-Jacobi-Bellman (HJB) equations as well as an optimality verification argument along the lines of Duffie, Gârleanu and Pedersen [2004] and Wang [2004]. In the last part, we show that the non-linear HJB equations admit a linear approximation when the coefficient of constant risk aversion is close to zero.

C.1 Investor's Problem

We fix probability space $(\Omega, \mathcal{F}, \mathcal{P})$, as well as a filtration $\{\mathcal{F}_t, t \geq 0\}$ satisfying the usual conditions (see Protter [1990]). An investor (low-valuation, high-valuation, or arbitrageur) can be of either one of finitely many types that we denote by $\tau \in \mathcal{T}$. The set $\mathcal{T}$ of type is the product of feasible porfolio holdings and of income-dividend correlations. The arrival times of trading counterparties and of changes of income-dividend correlations are counted by some adapted counting process $N_t \in \mathbb{N}^K$ with constant intensity $\gamma \in \mathbb{R}_+^K$. An investor with initial type τ_0 and initial wealth W_0 chooses

a predictable $\mathcal{T}$-valued type process τ_t and an adapted consumption and wealth process (c_t, W_t) subject to the following feasibility conditions. First, the type τ_t must remain constant during the inter-arrival times of the counting process N_t. Second, when the investor is in state $\tau \in \mathcal{T}$ and when the process $N_t(k)$ jumps, the investor can choose among transitions $\tau' \in \mathcal{T}'(\tau, k) \subseteq \mathcal{T}$. For example, when a buyer $\overline{b}$ meets a seller of asset i, he can either stay a buyer or make a transition to the lender type $\overline{\ell}i$. The investor's wealth process evolves according to the SDE

$$dW_t = (rW_t + m(\tau_t) - c_t)\,dt + \sqrt{\sigma(\tau_t)^2 + \sigma_e^2}\,d\tilde{B}_t + \sum_{k=1}^{K} P(\tau_{t-}, \tau_t)\,dN_t(k),$$

where $\tilde{B}_t$ is some adapted standard Brownian motion and, for all $(\tau, \tau') \in \mathcal{T}^2$, $P(\tau, \tau')$ is the payoff of making a transition from type τ to type τ'. For example, the payoff of making a transition from type $\overline{b}$ to type $\overline{\ell}i$ is $P(\overline{b}, \overline{\ell}i) = -p_i$. In addition, the wealth process must satisfy

$$\lim_{T\to\infty} E\left[\exp(-\beta T - r\alpha W_T)\right] = 0 \tag{C.1}$$

$$E\left(\int_0^T \exp(-zW_t)\,dt\right) < \infty, \tag{C.2}$$

for all $T \geq 0$ and $z \in \{r\alpha, 2r\alpha\}$. These conditions will be satisfied by our candidate optimal control and allow us to complete the standard optimality verification argument. The investor problem is to choose admissible type, wealth, and consumption processes in order to maximize intertemporal utility

$$-E\left[\int_0^\infty \exp(-\beta t - \alpha c_t)\,dt\right].$$

C.2 Hamilton Jacobi Bellman Equations

We guess an optimal control by seeking a value function $J : \mathbb{R} \times \mathcal{T} \to \mathbb{R}$ solving the following system of Hamilton Jacobi Bellman equations (HJB)

$$0 = \sup\Big\{-\exp\big(-\alpha\, c(\tau)\big) + \mathcal{D}^{(c,\tau')}J(W,\tau) - \beta J(W,\tau)\Big\}, \tag{C.3}$$

for all $\tau \in \mathcal{T}$, with respect to policy functions $c : \mathcal{T} \to \mathbb{R}$ and $\tau' : \mathcal{T} \times \{1, \ldots, K\} \to \mathcal{T}$, subject to $\tau'(\tau, k) \in \mathcal{T}'(\tau, k)$, and where

$$\begin{aligned} \mathcal{D}^{(c,\tau')} J(W,\tau) &\equiv J_W(W,\tau)\big(rW - c(\tau) + m(\tau)\big) + \frac{1}{2}\left(\sigma(\tau)^2 + \sigma_e^2\right) J_{WW}(W,\tau) \\ &\quad + \sum_{k=1}^{K} \gamma(k) \Big(J\big(W + P(\tau, \tau'(\tau,k)), \tau'(\tau,k)\big) - J(W,\tau) \Big). \end{aligned} \tag{C.4}$$

We guess that there exists a solution of the form

$$J(W,\tau) = -\frac{1}{r} \exp\left(-A(W + V(\tau)) + \frac{r - \beta + A^2\sigma_e^2/2}{r} \right), \tag{C.5}$$

where $A \equiv r\alpha$. Substituting our guess in (C.3) and maximizing with respect to consumption, we find that there exists a solution of the form (C.5) if an only if $V \in \mathbb{R}^{\mathcal{T}}$ solves

$$0 = -rV(\tau) + m(\tau) - \frac{A}{2}\sigma(\tau)^2 + \sum_{k=1}^{K} \gamma(k) \max_{\tau'(\tau,k) \in \mathcal{T}'(\tau,k)} \frac{1 - e^{-A\left(V(\tau'(\tau,k)) - V(\tau) + P(\tau,\tau'(\tau,k))\right)}}{A}, \tag{C.6}$$

for all $\tau \in \mathcal{T}$. The consumption maximizing (C.3) given $V(\tau)$ is

$$c(\tau) = r(W + V(\tau)) - \frac{r - \beta + A^2\sigma_e^2/2}{A}. \tag{C.7}$$

C.3 Optimality Verification

In this section we outline an optimality verification argument that closely follows Duffie, Gârleanu and Pedersen [2004] and Wang [2004]. Let's suppose that some V solves the system (C.6) and that the maximum is achieved by some policy function $\tau'(\,\cdot\,,\,\cdot\,)$. We verify that the investor's problem is solved by the type process τ_t^* that is generated recursively by the policy function $\tau'(\,\cdot\,,\,\cdot\,)$, together with the consumption and wealth processes

$$\begin{aligned} c_t^* &= r(W_t + V(\tau_t^*)) - \frac{r - \beta + A^2\sigma_e^2/2}{A} \\ dW_t^* &= \left(-rV(\tau_t^*) + \frac{r - \beta + A^2\sigma_e^2/2}{A} + m(\tau_t^*) \right) dt + \sqrt{\sigma(\tau_t^*) + \sigma_e^2}\, d\tilde{B}_t + \sum_{k=1}^{K} \gamma(k) P(\tau_{t-}^*, \tau_t^*)\, dN_t(k). \end{aligned}$$

The type process is feasible by construction. We only need to check conditions (C.1) and (C.2). Condition (C.2) holds because, for all $z \in \mathbb{R}$, the process $\exp(-zW_t)$ is a Geometric Brownian Motion, with state dependent and bounded drift, bounded volatility, and bounded jumps. One checks the transversality condition (C.1) by showing that, for each $T \geq 0$, $E\left[J(W_T^*, \tau_T^*)\right] = e^{(\beta-r)T} E\left[J(W_0, \tau_0)\right]$, following exactly the same steps as in Duffie, Gârleanu and Pedersen [2004]. Let's now consider any feasible type, consumption and wealth processes (τ_t, c_t, W_t). By Ito's Lemma

$$e^{-\beta t} J(W_t, \tau_t) = J(W_0, \tau_0) + \int_0^T e^{-\beta t} \left(\mathcal{D}J - \beta J\right) \, dt + \int_0^T e^{-\beta t} J_W(W_t, \tau_t) \sqrt{\sigma^2(\tau_t) + \sigma_e^2} dB_t$$
$$+ \sum_{k=1}^{K} \int_0^T e^{-\beta t} \big(J(W_{t-} + P(\tau_{t-}, \tau_t)) - J(W_{t-}, \tau_{t-})\big) \big(dN_t(k) - \gamma(k)\, dt\big).$$

The regularity condition (C.2) implies that the last two integral terms are martingales.[38] On the other hand, the HJB equations imply that $\mathcal{D}J - \beta J \leq \exp(-\alpha c_t)$. Replacing this inequality into the previous equation and taking expectations on both sides, we find

$$-E\left[\int_0^T \exp(-\alpha c_t - \beta t)\, dt\right] + E\left[e^{-\beta T} J(W_T, \tau_T)\right] \leq J(W_0, \tau_0) \tag{C.8}$$

with an equality for (τ_t^*, c_t^*, W_t^*). Then, letting T go to infinity in (C.8) and using the transversality condition (C.1), we find that the investor intertemporal utility is less or equal than $J(W_0, \tau_0)$ with an equality for (τ_t^*, c_t^*, W_t^*), establishing optimality.

C.4 First-Order Approximation

Let's assume that $x(\tau) \equiv A\sigma(\tau)$ does not depend on A. We study the family of HJB equations indexed by $A \in \mathbb{R}$:

$$0 = -rV(\tau) + m(\tau) - x(\tau) + \sum_{k=1}^{K} \gamma(k) \max_{\tau'(\tau,k) \in \mathcal{T}'(\tau,k)} H\Big(A\,,\, V\big(\tau'(\tau,k)\big) + P\big(\tau, \tau'(\tau,k)\big) - V(\tau)\Big), \tag{C.9}$$

for all $\tau \in \mathcal{T}$, where $H(x,y) \equiv (1 - e^{-xy})/x$ if $x > 0$ and $H(0,y) = 0$. Because the function H has power series expansion $\sum_{n=1}^{\infty} (-1)^n (x^{n-1} y^n)/n!$ it is infinitely continuously differentiable for all

[38] For the stochastic integral with respect to the Brownian motion, see Chapter 3 of Karatzas and Shreve [1991]. For the integral with respect to the compensated Poisson process, see Brémaud [1981], Theorem II-T8.

$(x, y) \in \mathbb{R}_+ \times \mathbb{R}$. We first show that the system (C.9) of HJB equation has a unique solution at $A = 0$. To see why this is the case one rewrite the system as

$$V(\tau) = \frac{m(\tau) - x(\tau)}{r + \sum_{k=1}^{K} \gamma(k)} + \sum_{k=1}^{K} \frac{\gamma(k)}{r + \sum_{j=1}^{K} \gamma(j)} \max_{\tau'(\tau,k) \in \mathcal{T}'(\tau,k)} \Big(V\big(\tau'(\tau, k)\big) + P\big(\tau, \tau'(\tau, k)\big) \Big), \quad \text{(C.10)}$$

for all $\tau \in \mathcal{T}$. Equation (C.10) defines a mapping from $\mathbb{R}^K$ to itself. The Blackwell sufficient condition (Stokey and Lucas [1989] page 54) implies that this mapping is a contraction, with modulus $(\gamma(1)+\ldots+\gamma(K))/(r+\gamma(1)+\ldots+\gamma(K)) < 1$. Therefore, an application of the Contraction Mapping Theorem (Stokey and Lucas [1989] page 54) shows that the system (C.10) has a unique solution that we denote V^0. We let $\tau^0(\,\cdot\,,\,\cdot\,)$ be a policy function achieving the maximum. Given $\tau^0(\,\cdot\,,\,\cdot\,)$, V^0 solves a system of linear equations that is invertible (because it can also be viewed as a contraction). We show

Proposition 9 (First-Order Approximation.) *If, for all $\tau \in \mathcal{T}$, $\tau^0(\tau, \cdot\,)$ is the unique maximizer of (C.10) given V^0, then there exists a neighborhood $N_1 \subseteq \mathbb{R}_+$ of zero, a neighborhood $N_2 \subseteq \mathbb{R}^K$ of V^0, and an infinitely continuously differentiable function $\phi : N_1 \to N_2$ such that, for all $A \in N_1$, V is a solution of the system (C.9) of HJB equations if and only if $V = \phi(A)$. Moreover, for all $A \in N_1$, for all $\tau \in \mathcal{T}$, $\tau^0(\tau, \cdot\,)$ is the unique maximizer of (C.9) given $V = \phi(A)$.*

We fix $\tau^0(\,\cdot\,,\,\cdot\,)$ and consider the system of equations

$$G(\tau, A, V) = -rV(\tau) + m(\tau) - x(\tau) + \sum_{k=1}^{K} \gamma(k) H\Big(A, V\big(\tau^0(\tau, k)\big) + P\big(\tau, \tau^0(\tau, k)\big) - V(\tau) \Big) = 0, \quad \text{(C.11)}$$

for all $\tau \in \mathcal{T}$. The function G is is infinitely continuously differentiable and its Jacobian at $(A, V) = (0, V^0)$ is invertible. Therefore an application of the Implicit Function Theorem (see Taylor and Mann [1983]) provides neighborhoods $\tilde{N}_1 \in \mathbb{R}_+$ and $\tilde{N}_2 \in \mathbb{R}^K$, and an infinitely continuously differentiable function ϕ such that, for all $A \in N_1$, $H(A, V) = 0$ if and only if $V = \phi(A)$. Because $\tau^0(\tau, \cdot\,)$ is the unique maximizer of (C.10), we know that for all feasible policy function τ', $\tau'(\tau, \cdot\,) \neq \tau^0(\tau, \cdot\,)$ implies

$$0 > -rV^0(\tau) + m(\tau) - x(\tau) + \sum_{k=1}^{K} \gamma(k) H\Big(0, V^0\big(\tau'(\tau, k)\big) + P\big(\tau, \tau'(\tau, k)\big) - V^0(\tau) \Big). \quad \text{(C.12)}$$

By continuity of H and ϕ, these strict inequalities hold in some neighborhood N_1 of zero. Therefore, for all $A \in N_1$, $V = \phi(A)$ is a solution of the system (C.9) of HJB equations, and τ^0 achieves the maximum.

D Utilities and Prices

We start by deriving the equations for the types' utilities and the lending fee. To do so, we need to expand the set of types, characterizing a high-valuation non-searcher by the state of his borrower. Depending on whether that agent is a seller $\underline{s}i$, non-searcher $\underline{n}i$, or buyer $\underline{b}i$, we denote the high-valuation non-searcher by $\overline{n}\underline{s}i$, $\overline{n}\underline{n}i$, and $\overline{n}\underline{b}i$, respectively. This ensures that transitions across types are Markovian.

Lender $\overline{\ell}i$: The equation is

$$rV_{\overline{\ell}i} = \delta + \overline{x} - y + \overline{\kappa}(V_{\overline{s}i} - V_{\overline{\ell}i}) + \nu_i \mu_{\underline{bo}}(V_{\overline{n}\underline{s}i} - V_{\overline{\ell}i}). \tag{D.1}$$

The flow benefit is the certainty equivalent $C(\overline{\rho}, 1) = \delta + \overline{x} - y$ of holding one share. The transitions are (i) revert to average valuation at rate $\overline{\kappa}$ and become a seller $\overline{s}i$, and (ii) meet a willing borrower at rate $\nu_i \mu_{\underline{bo}}$, lend the asset, and become of type $\overline{n}\underline{s}i$.

High-valuation non-searcher $\overline{n}\underline{s}i$: The equation is

$$rV_{\overline{n}\underline{s}i} = \delta + \overline{x} - y + w_i + \overline{\kappa}(V_{\overline{s}i} - V_{\overline{n}\underline{s}i}) + \underline{\kappa}(V_{\overline{\ell}i} - V_{\overline{n}\underline{s}i}) + \lambda \mu_{bi}(V_{\overline{n}\underline{n}i} - V_{\overline{n}\underline{s}i}). \tag{D.2}$$

The flow benefits are the certainty equivalent $C(\overline{\rho}, 1)$ of holding one share, and the lending fee w_i. The transitions are (i) revert to average valuation at rate $\overline{\kappa}$ and become a seller $\overline{s}i$, (ii) agent $\underline{s}i$ reverts to average valuation at rate $\underline{\kappa}$ and returns the asset, in which case agent $\overline{n}\underline{s}i$ becomes a lender $\overline{\ell}i$, and (iii) agent $\underline{s}i$ meets a seller at rate $\lambda \mu_{bi}$, in which case agent $\overline{n}\underline{s}i$ becomes of type $\overline{n}\underline{n}i$.

High-valuation non-searcher $\overline{n}\underline{n}i$: The equation is

$$rV_{\overline{n}\underline{n}i} = \delta + \overline{x} - y + w_i + \overline{\kappa}(C_i - V_{\overline{n}\underline{n}i}) + \underline{\kappa}(V_{\overline{n}\underline{b}i} - V_{\overline{n}\underline{n}i}). \tag{D.3}$$

The flow benefits are as for type $\overline{n}\underline{s}i$. The transitions are (i) revert to average valuation at rate $\overline{\kappa}$ and seize the collateral C_i, and (ii) agent $\underline{n}i$ reverts to average valuation at rate $\underline{\kappa}$, in which case agent $\overline{n}\underline{n}i$ becomes of type $\overline{n}\underline{b}i$.

High-valuation non-searcher $\overline{n}\underline{b}i$: The equation is

$$rV_{\overline{n}\underline{b}i} = \delta + \overline{x} - y + w_i + \overline{\kappa}(C_i - V_{\overline{n}\underline{b}i}) + \lambda \mu_{si}(V_{\overline{\ell}i} - V_{\overline{n}\underline{b}i}). \tag{D.4}$$

The flow benefits are as for type $\overline{n}\underline{s}i$. The transitions are (i) revert to average valuation at rate $\overline{\kappa}$ and seize the collateral C_i, and (ii) agent $\underline{b}i$ meets a seller at rate $\lambda\mu_{si}$ and returns the asset, in which case agent $\overline{n}\underline{b}i$ becomes a lender $\overline{\ell}i$.

Seller $\overline{s}i$: The equation is

$$rV_{\overline{s}i} = \delta - y + \lambda\mu_{bi}(p_i - V_{\overline{s}i}). \tag{D.5}$$

The flow benefit is the certainty equivalent $C(0,1) = \delta - y$ of holding one share. The only transition is to meet a buyer at rate $\lambda\mu_{bi}$, sell at price p_i, and exit the market.

Borrower $\underline{bo}$: The equation is

$$rV_{\underline{bo}} = -\underline{\kappa}V_{\underline{bo}} + \sum_{i=1}^{2} \nu_i \mu_{\overline{\ell}i}(V_{\underline{s}i} - V_{\underline{bo}}). \tag{D.6}$$

The flow benefits are zero. The transitions are (i) revert to average valuation at rate $\underline{\kappa}$ and exit the market, and (ii) borrow asset i at rate $\nu_i\mu_{\overline{\ell}i}$ and become a seller $\underline{s}i$.

Low-valuation seller $\underline{s}i$: The equation is

$$rV_{\underline{s}i} = -w_i + \overline{\kappa}(V_{\underline{bo}} - V_{\underline{s}i}) - \underline{\kappa}V_{\underline{s}i} + \lambda\mu_{bi}(V_{\underline{n}i} + p_i - V_{\underline{s}i}). \tag{D.7}$$

The flow benefit is paying the lending fee w_i. The transitions are (i) being asked to deliver (because the high-valuation agent reverts to average valuation at rate $\overline{\kappa}$) and become a borrower, (ii) revert to average valuation at rate $\underline{\kappa}$ and exit the market, and (iii) meet a buyer of asset i at rate $\lambda\mu_{bi}$, sell at price p_i, and become a low-valuation non-searcher $\underline{n}i$.

Low-valuation non-searcher $\underline{n}i$: The equation is

$$rV_{\underline{n}i} = -\delta + \underline{x} - y - w_i + \overline{\kappa}(V_{\underline{bo}} - C_i - V_{\underline{n}i}) + \underline{\kappa}(V_{\underline{b}i} - V_{\underline{n}i}). \tag{D.8}$$

The flow benefits are the certainty equivalent $C(\underline{\rho}, -1) = -\delta + \underline{x} - y$ of shorting one share, and paying the lending fee w_i. The transitions are (i) being asked to deliver at rate $\overline{\kappa}$, lose the collateral C_i, and become a borrower, and (ii) revert to average valuation at rate $\underline{\kappa}$ and become a buyer $\underline{b}i$.

Buyer $\underline{b}i$: The equation is

$$rV_{\underline{b}i} = -\delta - y - w_i + \overline{\kappa}(-C_i - V_{\underline{b}i}) + \lambda\mu_{si}(-p_i - V_{\underline{b}i}). \tag{D.9}$$

The flow benefits are the certainty equivalent $C(0,-1) = -\delta - y$ of shorting one share, and paying the lending fee w_i. The transitions are (i) being asked to deliver at rate $\overline{\kappa}$, lose the collateral C_i, and exit the market, and (ii) meet a seller at rate $\lambda\mu_{si}$, buy at price p_i, and exit the market.

Price p_i: Using Equation (14) and the definitions of $\Delta_{\overline{b}}$ and $\Delta_{\overline{s}i}$, we find

$$V_{\overline{\ell}i} - p_i - V_{\overline{b}} = \phi(V_{\overline{\ell}i} - V_{\overline{b}} - V_{\overline{s}i}). \tag{D.10}$$

The LHS is equal to the trading surplus of the marginal buyer $\overline{b}$, and the RHS is equal to the overall surplus (marginal buyer plus marginal seller) times the bargaining power ϕ.

Lending fee w_i: The counterpart of Equation (D.10) is

$$V_{\overline{n}\underline{s}i} - V_{\overline{\ell}i} = \theta(V_{\overline{n}\underline{s}i} + V_{\underline{s}i} - V_{\overline{\ell}i} - V_{\underline{b}o}) \equiv \theta\Sigma_i, \tag{D.11}$$

because the trading surplus of a lender $\overline{\ell}i$ is $V_{\overline{n}\underline{s}i} - V_{\overline{\ell}i}$, and the overall surplus is the sum of $V_{\overline{n}\underline{s}i} - V_{\overline{\ell}i}$ plus the borrower surplus $V_{\underline{s}i} - V_{\underline{b}o}$.

Using Equations (12) and (D.1)-(D.11), we will compute the lending fee w_i and the price p_i as a function of the short-selling surplus Σ_i. We will then derive a linear system for Σ_1 and Σ_2.

D.1 Lending Fee

Subtracting Equation (D.1) from (D.2), we find

$$(r + \overline{\kappa} + \underline{\kappa} + \nu_i\mu_{\underline{b}o})(V_{\overline{n}\underline{s}i} - V_{\overline{\ell}i}) = w_i + \lambda\mu_{bi}(V_{\overline{n}\underline{n}i} - V_{\overline{n}\underline{s}i}), \tag{D.12}$$

subtracting Equation (D.2) from (D.3), we find

$$(r + \overline{\kappa} + \underline{\kappa} + \lambda\mu_{bi})(V_{\overline{n}\underline{n}i} - V_{\overline{n}\underline{s}i}) = \overline{\kappa}(C_i - V_{\overline{s}i}) + \underline{\kappa}(V_{\overline{n}\underline{b}i} - V_{\overline{\ell}i}), \tag{D.13}$$

and subtracting Equation (D.3) from (D.4), we find

$$(r + \overline{\kappa} + \underline{\kappa})(V_{\overline{n}\underline{b}i} - V_{\overline{n}\underline{n}i}) = \lambda\mu_{si}(V_{\overline{\ell}i} - V_{\overline{n}\underline{b}i}). \tag{D.14}$$

Equations (D.13) and (D.14) imply that

$$V_{\overline{n}\underline{b}i} - V_{\overline{n}\underline{s}i} = \frac{\overline{\kappa}}{r + \overline{\kappa} + \underline{\kappa} + \lambda\mu_{bi}}(C_i - V_{\overline{s}i}) + \frac{\underline{\kappa}(r + \overline{\kappa} + \underline{\kappa}) - \lambda\mu_{si}(r + \overline{\kappa} + \underline{\kappa} + \lambda\mu_{bi})}{(r + \overline{\kappa} + \underline{\kappa})(r + \overline{\kappa} + \underline{\kappa} + \lambda\mu_{bi})}(V_{\overline{n}\underline{b}i} - V_{\overline{\ell}i}).$$

Adding $V_{\overline{n}\underline{s}i} - V_{\overline{\ell}i}$ on both sides and solving for $V_{\overline{n}\underline{b}i} - V_{\overline{\ell}i}$, we find

$$V_{\overline{n}\underline{b}i} - V_{\overline{\ell}i} = \frac{(r+\overline{\kappa}+\underline{\kappa})(r+\overline{\kappa}+\underline{\kappa}+\lambda\mu_{bi})}{(r+\overline{\kappa}+\underline{\kappa})(r+\overline{\kappa}+\lambda\mu_{bi})+\lambda\mu_{si}(r+\overline{\kappa}+\underline{\kappa}+\lambda\mu_{bi})}\left[\frac{\overline{\kappa}(C_i - V_{\overline{s}i})}{r+\overline{\kappa}+\underline{\kappa}+\lambda\mu_{bi}} + V_{\overline{n}\underline{s}i} - V_{\overline{\ell}i}\right]$$

Substituting $V_{\overline{n}\underline{b}i} - V_{\overline{\ell}i}$ from this equation into (D.13), we find

$$V_{\overline{n}\underline{n}i} - V_{\overline{n}\underline{s}i} = \frac{\overline{\kappa}(r+\overline{\kappa}+\underline{\kappa}+\lambda\mu_{si})(C_i - V_{\overline{s}i}) + \underline{\kappa}(r+\overline{\kappa}+\underline{\kappa})(V_{\overline{n}\underline{s}i} - V_{\overline{\ell}i})}{(r+\overline{\kappa}+\underline{\kappa})(r+\overline{\kappa}+\lambda\mu_{bi})+\lambda\mu_{si}(r+\overline{\kappa}+\underline{\kappa}+\lambda\mu_{bi})}.$$

Substituting $V_{\overline{n}\underline{n}i} - V_{\overline{n}\underline{s}i}$ from this equation into (D.12), and using $V_{\overline{n}\underline{s}i} - V_{\overline{\ell}i} = \theta\Sigma_i$ (i.e., Equation (D.11)), we can determine the lending fee as a function of the short-selling surplus:

$$\begin{aligned}
&\left[r+\overline{\kappa}+\underline{\kappa}\frac{(r+\overline{\kappa})(r+\overline{\kappa}+\underline{\kappa}+\lambda\mu_{si})+\lambda\mu_{si}(\underline{\kappa}+\lambda\mu_{bi})}{(r+\overline{\kappa}+\lambda\mu_{bi})(r+\overline{\kappa}+\underline{\kappa})+\lambda\mu_{si}(r+\overline{\kappa}+\underline{\kappa}+\lambda\mu_{bi})} + \nu_i\mu_{\underline{b}o}\right]\theta\Sigma_i \\
= \quad & w_i + \frac{\overline{\kappa}\lambda\mu_{bi}(r+\overline{\kappa}+\underline{\kappa}+\lambda\mu_{si})}{(r+\overline{\kappa}+\underline{\kappa})(r+\overline{\kappa}+\lambda\mu_{bi})+\lambda\mu_{si}(r+\overline{\kappa}+\underline{\kappa}+\lambda\mu_{bi})}(C_i - V_{\overline{s}i}).
\end{aligned} \tag{D.15}$$

D.2 Price

Equation (D.5) implies that

$$V_{\overline{s}i} - p_i = \frac{\delta - y - rp_i}{r+\lambda\mu_{bi}}. \tag{D.16}$$

Subtracting rp_i from both sides of Equation (D.1), and using (D.11) and (D.16), we find

$$V_{\overline{\ell}i} - p_i = \frac{1}{r+\overline{\kappa}}\left[\delta + \overline{x} - y - rp_i + \nu_i\mu_{\underline{b}o}\theta\Sigma_i + \overline{\kappa}\frac{\delta - y - rp_i}{r+\lambda\mu_{bi}}\right]. \tag{D.17}$$

Substituting (D.16) and (D.17) into (D.10), we find

$$\delta - y - rp_i + \frac{(1-\phi)(r+\lambda\mu_{bi})}{r+\overline{\kappa}+(1-\phi)\lambda\mu_{bi}}\left[\overline{x} + \nu_i\mu_{\underline{b}o}\theta\Sigma_i - (r+\overline{\kappa})V_{\overline{b}}\right] = 0. \tag{D.18}$$

Substituting $d - y - rp_i$ from Equation (D.18) into (D.17), we find

$$V_{\overline{\ell}i} - p_i = \frac{\phi(\overline{x} + \nu_i\mu_{\underline{b}o}\theta\Sigma_i) + (1-\phi)(r+\overline{\kappa}+\lambda\mu_{bi})V_{\overline{b}}}{r+\overline{\kappa}+(1-\phi)\lambda\mu_{bi}}.$$

Substituting $V_{\overline{\ell}i} - p_i$ from this equation into (12) and solving for $V_{\overline{b}}$, we find

$$V_{\overline{b}} = \frac{\phi \sum_{j=1}^{2} \frac{\lambda \mu_{sj}}{r+\overline{\kappa}+(1-\phi)\lambda\mu_{bj}} (\overline{x} + \nu_j \mu_{\underline{bo}} \theta \Sigma_j)}{(r+\overline{\kappa}) \left[1 + \phi \sum_{j=1}^{2} \frac{\lambda \mu_{sj}}{r+\overline{\kappa}+(1-\phi)\lambda\mu_{bj}}\right]}.$$

Substituting $V_{\overline{b}}$ from this equation into (D.18), we can determine the price as a function of the short-selling surplus:

$$p_i = \frac{\delta - y}{r} + \frac{(1-\phi)(r+\lambda\mu_{bi})}{r\left[r+\overline{\kappa}+(1-\phi)\lambda\mu_{bi}\right]} \left[\overline{x} + \nu_i \mu_{\underline{bo}} \theta \Sigma_i - \frac{\phi \sum_{j=1}^{2} \frac{\lambda \mu_{sj}}{r+\overline{\kappa}+(1-\phi)\lambda\mu_{bj}} (\overline{x} + \nu_j \mu_{\underline{bo}} \theta \Sigma_j)}{1 + \phi \sum_{j=1}^{2} \frac{\lambda \mu_{sj}}{r+\overline{\kappa}+(1-\phi)\lambda\mu_{bj}}}\right]. \tag{D.19}$$

D.3 Short-Selling Surplus

Adding Equations (D.2) and (D.7), and subtracting Equations (D.6) and (D.1), we find

$$(r+\overline{\kappa}+\underline{\kappa}+\nu_i\mu_{\underline{bo}}\theta)\Sigma_i + \sum_{j=1}^{2} \nu_j \mu_{\overline{\ell}j}(1-\theta)\Sigma_j = \lambda\mu_{bi}(V_{\overline{n}\underline{n}i} + V_{\underline{n}i} + p_i - V_{\overline{n}\underline{s}i} - V_{\underline{s}i}). \tag{D.20}$$

Adding Equations (D.3), (D.8), and $rp_i = rp_i$, and subtracting Equations (D.2) and (D.7), we find

$$(r+\overline{\kappa}+\underline{\kappa}+\lambda\mu_{bi})(V_{\overline{n}\underline{n}i}+V_{\underline{n}i}+p_i-V_{\overline{n}\underline{s}i}-V_{\underline{s}i}) = rp_i-\delta+\underline{x}-y+\overline{\kappa}(p_i-V_{\overline{s}i})+\underline{\kappa}(V_{\overline{n}\underline{b}i}+V_{\underline{b}i}+p_i-V_{\overline{\ell}i}). \tag{D.21}$$

Adding Equations (D.4), (D.9), and $rp_i = rp_i$, and subtracting Equation (D.1), we find

$$(r+\overline{\kappa}+\lambda\mu_{si})(V_{\overline{n}\underline{b}i} + V_{\underline{b}i} + p_i - V_{\overline{\ell}i}) = rp_i - \delta - y + \overline{\kappa}(p_i - V_{\overline{s}i}) - \nu_i\mu_{\underline{bo}}\theta\Sigma_i. \tag{D.22}$$

Substituting $V_{\overline{n}\underline{b}i} + V_{\underline{b}i} + p_i - V_{\overline{\ell}i}$ from Equation (D.22) into (D.21), and then substituting $V_{\overline{n}\underline{n}i} + V_{\underline{n}i} + p_i - V_{\overline{n}\underline{s}i} - V_{\underline{s}i}$ from Equation (D.21) into (D.20), we find

$$\begin{aligned} &\left[r+\overline{\kappa}+\underline{\kappa}+\nu_i\mu_{\underline{bo}}\theta\left[1+\frac{\lambda\mu_{bi}\underline{\kappa}}{(r+\overline{\kappa}+\underline{\kappa}+\lambda\mu_{bi})(r+\overline{\kappa}+\lambda\mu_{si})}\right]\right]\Sigma_i + \sum_{j=1}^{2}\nu_j\mu_{\overline{\ell}j}(1-\theta)\Sigma_j \\ = \;& \frac{\lambda\mu_{bi}}{r+\overline{\kappa}+\underline{\kappa}+\lambda\mu_{bi}}\left[\underline{x} + \frac{r+\overline{\kappa}+\underline{\kappa}+\lambda\mu_{si}}{r+\overline{\kappa}+\lambda\mu_{si}}\left[rp_i-\delta-y+\overline{\kappa}(p_i-V_{\overline{s}i})\right]\right]. \end{aligned} \tag{D.23}$$

To derive an equation involving only Σ_1 and Σ_2, we need to eliminate the price p_i. We have

$$\begin{aligned} & rp_i - \delta - y + \overline{\kappa}(p_i - V_{\overline{s}i}) \\ = \; & -2y + rp_i - \delta + y + \overline{\kappa}\frac{rp_i - \delta + y}{r + \lambda\mu_{bi}} \\ = \; & -2y + \frac{(1-\phi)(r+\overline{\kappa}+\lambda\mu_{bi})}{r+\overline{\kappa}+(1-\phi)\lambda\mu_{bi}}\left[\overline{x} + \nu_i\mu_{\underline{bo}}\theta\Sigma_i - \frac{\phi\sum_{j=1}^{2}\frac{\lambda\mu_{sj}}{r+\overline{\kappa}+(1-\phi)\lambda\mu_{bj}}(\overline{x}+\nu_j\mu_{\underline{bo}}\theta\Sigma_j)}{1+\phi\sum_{j=1}^{2}\frac{\lambda\mu_{sj}}{r+\overline{\kappa}+(1-\phi)\lambda\mu_{bj}}}\right], \end{aligned}$$

where the first step follows from Equation (D.16) and the second from (D.19). Plugging back into Equation (D.23), we can write it as

$$a_i\Sigma_i + \sum_{j=1}^{2} f_j\Sigma_j + b_i\sum_{j=1}^{2} g_j\Sigma_j = c_i, \tag{D.24}$$

where

$$a_i = r+\overline{\kappa}+\underline{\kappa}+\nu_i\mu_{\underline{bo}}\theta\left[\frac{r+\overline{\kappa}+\underline{\kappa}}{r+\overline{\kappa}+\underline{\kappa}+\lambda\mu_{bi}} + \frac{\phi(r+\overline{\kappa})\lambda\mu_{bi}(r+\overline{\kappa}+\underline{\kappa}+\lambda\mu_{si})}{(r+\overline{\kappa}+\underline{\kappa}+\lambda\mu_{bi})(r+\overline{\kappa}+\lambda\mu_{si})[r+\overline{\kappa}+(1-\phi)\lambda\mu_{bi}]}\right],$$

$$f_i = \nu_i\mu_{\overline{l}i}(1-\theta),$$

$$b_i = \frac{(1-\phi)\lambda\mu_{bi}(r+\overline{\kappa}+\underline{\kappa}+\lambda\mu_{si})(r+\overline{\kappa}+\lambda\mu_{bi})}{(r+\overline{\kappa}+\underline{\kappa}+\lambda\mu_{bi})(r+\overline{\kappa}+\lambda\mu_{si})\,[r+\overline{\kappa}+\lambda(1-\phi)\lambda\mu_{bi}]},$$

$$g_i = \phi\nu_i\mu_{\underline{bo}}\theta\frac{\frac{\lambda\mu_{si}}{r+\overline{\kappa}+(1-\phi)\lambda\mu_{bi}}}{1+\phi\sum_{j=1}^{2}\frac{\lambda\mu_{sj}}{r+\overline{\kappa}+(1-\phi)\lambda\mu_{bj}}},$$

$$c_i = \frac{\lambda\mu_{bi}}{r+\overline{\kappa}+\underline{\kappa}+\lambda\mu_{bi}}\left[\underline{x} - \frac{r+\overline{\kappa}+\underline{\kappa}+\lambda\mu_{si}}{r+\overline{\kappa}+\lambda\mu_{si}}\left[2y - \frac{(1-\phi)(r+\overline{\kappa}+\lambda\mu_{bi})}{r+\overline{\kappa}+(1-\phi)\lambda\mu_{bi}}\frac{\overline{x}}{1+\phi\sum_{j=1}^{2}\frac{\lambda\mu_{sj}}{r+\overline{\kappa}+(1-\phi)\lambda\mu_{sj}}}\right]\right].$$

The short-selling surpluses Σ_1 and Σ_2 are the solution to the linear system consisting of Equation (D.24) for $i \in \{1, 2\}$.

Note that the collateral C_i does not enter in Equation (D.24), and thus does not affect the short-selling surplus. It also does not affect the price, from Equation (D.19). It affects only the lending fee because when lenders can seize more collateral they accept a lower fee. From now on (and as stated in Footnote 19), we set the collateral equal to the utility of a seller $\overline{s}i$, i.e.,

$$C_i = V_{\overline{s}i}. \tag{D.25}$$

E Search Equilibrium

In this section we prove Propositions 2-8.

Proof of Proposition 2: From Appendix B we know that given the short-selling decisions $\nu_1 = \nu_2 = \nu$, the types' measures are uniquely determined. From Appendix D we know that given any short-selling decisions and types' measures, the utilities, prices, and lending fees are uniquely determined. Therefore, what is left to show is (i) the short-selling surplus $\Sigma_1 = \Sigma_2$ is positive, (ii) buyers' and sellers' reservation values are ordered as in Equation (13), and (iii) agents' portfolio decisions are optimal. To show these results, we recall from Appendix B that when search frictions become small, i.e., λ goes to ∞ holding $n \equiv \nu/\lambda$ constant, μ_{bi} converges to m_b, $\mu_{\overline{\ell}i}$ converges to S, $\lambda\mu_{si}$ converges to g_s, and $\nu\mu_{\underline{bo}}$ converges to $g_{\underline{bo}}$.

We start by computing Σ_i, w_i, and p_i, thus proving Proposition 3. Equation (D.24) implies that when $\Sigma_1 = \Sigma_2 \equiv \Sigma$,

$$\Sigma = \frac{c}{a + 2(f + bg)},$$

where we suppress the asset subscripts from a, b, c, f, g because of symmetry. When search frictions become small, a and b converge to positive limits, c converges to

$$\underline{x} - \frac{r + \overline{\kappa} + \underline{\kappa} + g_s}{r + \overline{\kappa} + g_s}(2y - \overline{x}), \tag{E.1}$$

g converges to zero, and f converges to ∞, being asymptotically equal to $\nu S(1-\theta)$. Therefore, the surplus converges to zero, and its asymptotic behavior is as in Proposition 3.

Equations (D.15) and (D.25) imply that the lending fee is

$$w_i = \left[r + \overline{\kappa} + \underline{\kappa}\frac{(r + \overline{\kappa})(r + \overline{\kappa} + \underline{\kappa} + \lambda\mu_{si}) + \lambda\mu_{si}(\underline{\kappa} + \lambda\mu_{bi})}{(r + \overline{\kappa} + \lambda\mu_{bi})(r + \overline{\kappa} + \underline{\kappa}) + \lambda\mu_{si}(r + \overline{\kappa} + \underline{\kappa} + \lambda\mu_{bi})} + \nu_i\mu_{\underline{bo}}\right]\theta\Sigma_i.$$

Because the term in brackets converges to

$$r + \overline{\kappa} + \underline{\kappa}\frac{g_s}{r + \overline{\kappa} + \underline{\kappa} + g_s} + g_{\underline{bo}},$$

the lending fee converges to zero, and its asymptotic behavior is as in Proposition 3.

Equation (D.19) implies that the price is equal to

$$p_i = \frac{\delta - y}{r} + \frac{1}{r}\left[1 - \frac{\phi r + \overline{\kappa}}{(1-\phi)\lambda m_b} + o(1/\lambda)\right]\left[\overline{x} + g_{\underline{bo}}\theta\Sigma_i - \frac{2\phi g_s \overline{x}}{(1-\phi)\lambda m_b} + o(1/\lambda)\right].$$

Using this equation and the fact that Σ_i is in order $1/\lambda$, it is easy to check that the asymptotic behavior (i.e., order $1/\lambda$) of the price is as in Proposition 3.

To show that Σ_i is positive, we need to show that Equation (E.1) is positive. This follows because Equation (17) implies that

$$\underline{x} > 2y + \frac{\underline{\kappa}}{r + \overline{\kappa} + g_s}(2y - \overline{x}) > 2y - \overline{x} + \frac{\underline{\kappa}}{r + \overline{\kappa} + g_s}(2y - \overline{x}) = \frac{r + \overline{\kappa} + \underline{\kappa} + g_s}{r + \overline{\kappa} + g_s}(2y - \overline{x}). \quad \text{(E.2)}$$

We next show that reservation values are ordered as in Equation (13), i.e., $\Delta_{\underline{b}i} > \Delta_{\overline{b}}$ and $\Delta_{\overline{s}i} > \Delta_{\underline{s}i}$. For this, we need to compute $V_{\underline{b}i}$ and $V_{\underline{n}i} - V_{\underline{s}i}$. Adding Equations (D.9) and $rp_i = rp_i$, and using Equation (D.25), we find

$$V_{\underline{b}i} + p_i = \frac{rp_i - \delta - y - w_i + \overline{\kappa}(p_i - V_{\overline{s}i})}{r + \overline{\kappa} + \lambda\mu_{si}}. \quad \text{(E.3)}$$

Adding Equations (D.8) and $rp_i = rp_i$, and subtracting Equation (D.7), we similarly find

$$V_{\underline{n}i} + p_i - V_{\underline{s}i} = \frac{rp_i - \delta + \underline{x} - y + \underline{\kappa}(V_{\underline{b}i} + p_i) + \overline{\kappa}(p_i - V_{\overline{s}i})}{r + \overline{\kappa} + \underline{\kappa} + \lambda\mu_{bi}}. \quad \text{(E.4)}$$

Inequality $\Delta_{\underline{b}i} > \Delta_{\overline{b}}$ is equivalent to

$$\begin{aligned}
& -V_{\underline{b}i} - p_i > V_{\overline{\ell}i} - p_i - V_{\overline{b}} \\
\Leftrightarrow \quad & \frac{\delta + y - rp_i + w_i - \overline{\kappa}(p_i - V_{\overline{s}i})}{r + \overline{\kappa} + \lambda\mu_{si}} > \frac{\phi}{1-\phi}(p_i - V_{\overline{s}i}) \\
\Leftrightarrow \quad & \frac{\delta + y - rp_i + w_i - \overline{\kappa}\frac{rp_i - \delta + y}{r + \lambda\mu_{bi}}}{r + \overline{\kappa} + \lambda\mu_{si}} > \frac{\phi}{1-\phi}\frac{rp_i - \delta + y}{r + \lambda\mu_{bi}}
\end{aligned} \quad \text{(E.5)}$$

where the second step follows from Equations (D.10) and (E.3), and the third from Equation (D.16). Because rp_i converges to $\delta + \overline{x} - y$, and w_i converges to zero, the LHS of Equation (E.5) converges to $(2y - \overline{x})/(r + \overline{\kappa} + g_s)$, which is positive from Equation (17), while the RHS converges to zero.

Inequality $\Delta_{\overline{s}i} > \Delta_{\underline{s}i}$ is equivalent to

$$\begin{aligned} & V_{\underline{n}i} + p_i - V_{\underline{s}i} > p_i - V_{\overline{s}i} \\ \Leftrightarrow \quad & \frac{\underline{x} + \frac{r+\overline{\kappa}+\underline{\kappa}+\lambda\mu_{si}}{r+\underline{\kappa}+\lambda\mu_{si}}[rp_i - \delta - y + \overline{\kappa}(p_i - V_{\overline{s}i})] - \frac{\underline{\kappa}}{r+\overline{\kappa}+\lambda\mu_{si}} w_i}{r + \overline{\kappa} + \underline{\kappa} + \lambda\mu_{bi}} > \frac{rp_i - \delta + y}{r + \lambda\mu_{bi}}, \end{aligned}$$

where the second step follows from Equations (D.16), (E.3), and (E.4). When search frictions become small, this inequality holds if the limit of the numerator in the LHS exceeds that forss the RHS, i.e.,

$$\underline{x} - \frac{r + \overline{\kappa} + \underline{\kappa} + g_s}{r + \underline{\kappa} + g_s}(2y - \overline{x}) > \overline{x}.$$

This holds because of the first inequality in Equation (E.2).

We finally show that portfolio decisions are optimal. The flow benefit that an average-valuation agent can derive from a long position in asset i is bounded above by $\delta - y + w_i$, and the flow benefit for a short position is bounded above by $-\delta - y$. Therefore, an average-valuation agent finds it optimal to establish no position, or to unwind a previously established one, if $(\delta - y + w_i)/r < \min\{p_i, C_i\}$ and $(\delta + y)/r > p_i$. These conditions are satisfied for small frictions because p_i converges to $(\delta + \overline{x} - y)/r$, w_i converges to zero, $C_i - p_i$ converges to zero, and $2y > \overline{x}$.

A high-valuation agent finds it optimal to buy asset i if $V_{\overline{\ell}i} - p_i - V_{\overline{b}} \geq 0$. This condition is satisfied because

$$V_{\overline{\ell}i} - p_i - V_{\overline{b}} = \frac{\phi}{1-\phi}(p_i - V_{\overline{s}i}) = \frac{\phi}{1-\phi}\frac{rp_i - \delta + y}{r + \lambda\mu_{bi}} \sim \frac{\phi}{1-\phi}\frac{\overline{x}}{\lambda\mu_{bi}} \geq 0.$$

The agent then finds it optimal to lend the asset because $V_{\overline{n}\underline{s}i} - V_{\overline{\ell}i} = \theta\Sigma_i > 0$. Likewise, a low-valuation agent finds it optimal to borrow asset i because $V_{\underline{s}i} - V_{\underline{b}o} = (1-\theta)\Sigma_i > 0$, and to sell it because $V_{\underline{n}i} + p_i - V_{\underline{s}i} = p_i - \Delta_{\underline{s}i} > p_i - \Delta_{\overline{s}i} = p_i - V_{\overline{s}i} > 0$. Finally, an arbitrage portfolio is not profitable because the two assets trade at the same price and carry the same lending fee. ∎

Proof of Proposition 3: See the proof of Proposition 2. ∎

Proof of Proposition 4: We need to show that (i) the short-selling surplus Σ_1 is positive and Σ_2 is negative, (ii) buyers' and sellers' reservation values are ordered as in Equation (13), and (iii) agents' portfolio decisions are optimal. We recall from Appendix B that for small search frictions

and given the short-selling decisions $\nu_1 = \nu$ and $\nu_2 = 0$, μ_{bi} converges to $\hat{m}_{bi}$, $\mu_{\overline{\ell}i}$ converges to S, $\lambda\mu_{si}$ converges to $\hat{g}_{si}$, and $\nu\mu_{\underline{bo}}$ converges to $\hat{g}_{\underline{bo}}$.

We start by computing Σ_1, w_1, p_1, and p_2, thus proving Proposition 5. Equation (D.24) implies that when $\nu_2 = 0$,

$$\Sigma_1 = \frac{c_1}{a_1 + f_1 + b_1 g_1}.$$

When search frictions become small, c_1 converges to

$$\underline{x} - \frac{r + \overline{\kappa} + \underline{\kappa} + \hat{g}_{s1}}{r + \overline{\kappa} + \hat{g}_{s1}}(2y - \overline{x}), \tag{E.6}$$

and the dominant term in the denominator is $f_1 \sim \nu S(1-\theta)$. Therefore, the surplus converges to zero, and its asymptotic behavior is as in Proposition 5. To determine the asymptotic behavior of the lending fee and the price, we proceed as in the proof of Proposition 2.

To show that Σ_1 is positive, we need to show that Equation (E.6) is positive. This follows from Equation (E.2) and the fact that $\hat{g}_{s1} > g_s$, established in the proof of Proposition 6. To show that Σ_2 is negative, we note that from Equation (D.24),

$$\Sigma_2 = \frac{c_2 - (f_1 + b_2 g_1)\Sigma_1}{a_2} = \frac{c_2 - \frac{f_1 + b_2 g_1}{a_1 + f_1 + b_1 g_1} c_1}{a_2}.$$

When search frictions become small, the numerator converges to the same limit as $c_2 - c_1$. This limit is equal to

$$\left[\frac{r + \overline{\kappa} + \underline{\kappa} + \hat{g}_{s1}}{r + \overline{\kappa} + \hat{g}_{s1}} - \frac{r + \overline{\kappa} + \underline{\kappa} + \hat{g}_{s2}}{r + \overline{\kappa} + \hat{g}_{s2}}\right](2y - \overline{x}),$$

and is negative if $\hat{g}_{s1} > \hat{g}_{s2}$. Using Equations (B.31) and (B.33), we can write this inequality as

$$\frac{\overline{\kappa} S + \frac{\overline{\kappa} + \underline{\kappa}}{\underline{\kappa}} \underline{F}}{\hat{m}_{b1}} > \frac{\overline{\kappa} S}{\hat{m}_{\overline{b}}}. \tag{E.7}$$

Equations (B.33)-(B.35) imply that

$$\hat{m}_{\overline{b}} = \frac{\overline{F} - \overline{\kappa} S}{\overline{F} - \overline{\kappa} S + \underline{F}} \hat{m}_{b1}. \tag{E.8}$$

Using this equation, we can write Equation (E.7) as

$$\frac{\overline{\kappa}S + \frac{\overline{\kappa}+\kappa}{\underline{\kappa}}\underline{F}}{\overline{\kappa}S} > \frac{\overline{F} - \overline{\kappa}S + \underline{F}}{\overline{F} - \overline{\kappa}S}.$$

It is easy to check that this inequality holds because of Equation (16).

To show that $\Delta_{\underline{b}i} > \Delta_{\overline{b}}$ and $\Delta_{\overline{s}i} > \Delta_{\underline{s}i}$, we proceed as in the proof of Proposition 2. The only change is that the condition for $\Delta_{\overline{s}i} > \Delta_{\underline{s}i}$ now is

$$\underline{x} - \frac{r + \overline{\kappa} + \underline{\kappa} + \hat{g}_s}{r + \underline{\kappa} + \hat{g}_s}(2y - \overline{x}) > \overline{x}.$$

This inequality is implied by the first inequality in Equation (E.2) and the fact that $\hat{g}_{s1} > g_s$.

The arguments in the proof of Proposition 2 establish portfolio optimality for all agents except the arbitrageurs. Arbitrageurs could attempt to exploit the price differential in the asymmetric equilibrium by buying one asset and shorting the other. We next show that buying asset 2 and shorting asset 1 is unprofitable under

$$p_1 - p_2 < \frac{w_1}{r} + \frac{\overline{x}}{\lambda \hat{m}_{b1}} + \frac{\overline{\kappa x}}{r(\nu S + \lambda \hat{m}_{b2})}. \tag{E.9}$$

(which is implied by (25)), while buying asset 1 and shorting asset 2 is unprofitable under (26). We then show that Equations (25) and (26) are satisfied if ν/λ is in an interval (n_1, n_2).

Buy asset 2, short asset 1

Because trading opportunities arrive one at a time, an arbitrageur cannot set up the two legs of the position simultaneously. The arbitrageur can, for example, buy asset 2 first, then borrow asset 1, and then sell asset 1. Alternatively, he can borrow asset 1 first, then buy asset 2, and then sell asset 1. The final possibility, which is to sell asset 1 before buying asset 2 is suboptimal. Indeed, for small search frictions the time to meet a buyer converges to zero while the time to meet a seller does not. Therefore, the cost of being unhedged converges to zero only when asset 2 is bought before asset 1 is sold.

Suppose now that the arbitrage strategy is profitable. Because the payoff of the strategy is decreasing in asset 1's lending fee, there exists a fee $\overline{w}_1 > w_1$ for which the arbitrageur is indifferent between following the strategy and holding no position. If for this fee it is optimal to initiate the strategy by buying asset 2, the arbitrageur can be in three possible states:

(i) Long position in asset 2. State $n2$ with utility V_{n2}.

(ii) Long position in asset 2 and borrowed asset 1. State $s1n2$ with utility V_{s1n2}.

(iii) Long position in asset 2 and short in asset 1. State $n1n2$ with utility V_{n1n2}.

The utilities are characterized by the following flow-value equations:

$$rV_{n2} = \delta - y + \nu\mu_{\overline{\ell}1}(V_{s1n2} - V_{n2}) \tag{E.10}$$

$$rV_{s1n2} = \delta - y - \overline{w}_1 + \lambda\mu_{b1}(V_{n1n2} + p_1 - V_{s1n2}) + \overline{\kappa}(V_{n2} - V_{s1n2}) \tag{E.11}$$

$$rV_{n1n2} = -\overline{w}_1 + \overline{\kappa}(V_{n2} - C_1 - V_{n1n2}). \tag{E.12}$$

When in state $n2$, the arbitrageur receives the certainty equivalent $C(0,1) = \delta - y$ of holding one share, and can transit to state $s1n2$ by borrowing asset 1. When in state $s1n2$, the arbitrageur receives $\delta - y$, pays the lending fee $\overline{w}_1$, can transit to state $n1n2$ by selling asset 1, and can transit to state $n2$ if the lender calls for delivery. Finally, when in state $n1n2$, the arbitrageur is fully hedged, pays the lending fee, and can transit to state $n2$ if the lender calls for delivery. Solving Equations (E.10)-(E.12), we find

$$rV_{n2} = \delta - y + \frac{\nu\mu_{\overline{\ell}1}}{r + \overline{\kappa} + \nu\mu_{\overline{\ell}1}}\left[-\overline{w}_1 + \frac{\lambda\mu_{b1}}{r + \overline{\kappa} + \lambda\mu_{b1}}\left[rp_1 - \delta + y + \overline{\kappa}(p_1 - C_1)\right]\right].$$

The arbitrageur is indifferent between initiating the strategy and holding no position if V_{n2} is equal to p_2. Using this condition, and substituting C_1 from Equations (D.16) and (D.25), we find

$$\overline{w}_1 = \frac{\lambda\mu_{b2}}{r + \lambda\mu_{b2}}(rp_1 - \delta + y) - \frac{r + \overline{\kappa} + \nu\mu_{\overline{\ell}2}}{\nu\mu_{\overline{\ell}2}}(rp_2 - \delta - y).$$

For small search frictions, this equation becomes

$$\overline{w}_1 = r(p_1 - p_2) - \frac{r\overline{x}}{\lambda\hat{m}_{b1}} - \frac{(r + \overline{\kappa})\overline{x}}{\nu S},$$

and is inconsistent with Equation (E.9) since $w_1 < \overline{w}_1$.

Suppose instead that it is optimal to initiate the strategy by borrowing asset 1. The arbitrageur then starts from a state $s1$, in which he has borrowed asset 1 but holds no position in asset 2. The utility V_{s1} in this state is characterized by

$$rV_{s1} = -w_1 + \lambda\mu_{s2}(V_{s1n2} - p_2 - V_{s1}), \tag{E.13}$$

because the flow benefit is to pay the lending fee, and the transition is to state $s1n2$ by buying asset 1. The utility in states $s1n2$ and $n1n2$ is given by Equations (E.11) and (E.12), respectively. The utility in state $n2$, however, is given by

$$rV_{n2} = \delta - y + \nu\mu_{\overline{\ell}1}(V_{s1n2} - V_{n2}) + \lambda\mu_{b2}(p_2 - V_{n2}) \tag{E.14}$$

instead of (E.10). Indeed, since it it suboptimal to initiate the strategy by buying asset 2, buying that asset is dominated by holding no position. Therefore, if the arbitrageur finds himself with a long position in asset 2, he prefers to unwind it upon meeting a seller. Equations (E.11), (E.12), and (E.14) imply that

$$V_{s1n2} = \frac{\frac{r+\overline{\kappa}+\nu\mu_{\overline{\ell}1}+\lambda\mu_{b2}}{r+\nu\mu_{\overline{\ell}1}+\lambda\mu_{b2}}(\delta - y) + \frac{\overline{\kappa}\lambda\mu_{b2}}{r+\nu\mu_{\overline{\ell}1}+\lambda\mu_{b2}}p_2 - \overline{w}_1 + \frac{\lambda\mu_{b1}}{r+\overline{\kappa}+\lambda\mu_{b1}}\left[rp_1 - \delta + y + \overline{\kappa}(p_1 - C_1)\right]}{\frac{r(r+\overline{\kappa}+\nu\mu_{\overline{\ell}1}+\lambda\mu_{b2})+\overline{\kappa}\lambda\mu_{b2}}{r+\nu\mu_{\overline{\ell}1}+\lambda\mu_{b2}}}.$$

Plugging into Equation (E.13), and using Equations (D.16), (D.25), and the indifference condition which now is $V_{s1} = 0$, we find

$$\overline{w}_1 = \frac{\frac{\lambda\mu_{b1}}{r+\lambda\mu_{b1}}(rp_1 - \delta + y) - \frac{r+\overline{\kappa}+\nu\mu_{\overline{\ell}1}+\lambda\mu_{b2}}{r+\nu\mu_{\overline{\ell}1}+\lambda\mu_{b2}}(rp_2 - \delta + y)}{1 + \frac{r(r+\overline{\kappa}+\nu\mu_{\overline{\ell}1}+\lambda\mu_{b2})+\overline{\kappa}\lambda\mu_{b2}}{\lambda\mu_{s2}(r+\nu\mu_{\overline{\ell}1}+\lambda\mu_{b2})}}.$$

For small search frictions, this equation becomes

$$\overline{w}_1 = \frac{r(p_1 - p_2) - \frac{r\overline{x}}{\lambda\hat{m}_{b1}} - \frac{\overline{\kappa}\overline{x}}{\nu S+\lambda\hat{m}_{b2}}}{1 + \frac{r(nS+\hat{m}_{b2})+\overline{\kappa}\hat{m}_{b2}}{\hat{g}_{s2}(nS+\hat{m}_{b2})}},$$

and is inconsistent with Equation (E.9) since $w_1 < \overline{w}_1$.

Buy asset 1, short asset 2

We consider a "relaxed" problem where asset 1 can be bought instantly and asset 2 can be borrowed instantly at a lending fee of zero. Clearly, if the arbitrage strategy is unprofitable in the relaxed problem, it is also unprofitable when more frictions are present.

Suppose that the arbitrage strategy is profitable. Because the payoff of the strategy is increasing in asset 1's lending fee, there exists a fee $\overline{w}_1 < w_1$ for which the arbitrageur is indifferent between following the strategy and holding no position. When following the strategy, the arbitrageur is always in a state where he holds asset 1 and has borrowed asset 2, because these can be done instantly. If the arbitrageur has not sold asset 2, he can be in four possible states:

(i) Seeking to lend asset 1. State $\ell 1 s 2$ with utility $V_{\ell 1 s 2}$.

(ii) Lent asset 1 to an agent $\underline{s}1$. State $n\underline{s}1s2$ with utility $V_{n\underline{s}1s2}$.

(iii) Lent asset 1 to an agent $\underline{n}1$. State $n\underline{n}1s2$ with utility $V_{n\underline{n}1s2}$.

(iv) Lent asset 1 to an agent $\underline{b}1$. State $n\underline{b}1s2$ with utility $V_{n\underline{b}1s2}$.

If the arbitrageur has sold asset 2, he can be in the four corresponding states that we denote with $n2$ instead of $s2$.

For brevity, we skip the eight flow-value equations, but note that they have a simple solution. To each outcome concerning asset 1 ($\ell 1$, $n\underline{s}1$, $n\underline{n}1$, $n\underline{b}1$) and to each outcome concerning asset 2 ($s2$, $n2$), we can associate a separate utility that we denote by $\hat{V}$. We can then write the utility of a state (which is a "joint" outcome) as the sum of the two separate utilities. For example, the utility $V_{\ell 1 s 2}$ is equal to $\hat{V}_{\ell 1} + \hat{V}_{s2}$. This decomposition is possible because the outcomes concerning each asset evolve independently.

The utilities $\hat{V}_{\ell 1}$, $\hat{V}_{n\underline{s}1}$, $\hat{V}_{n\underline{n}1}$, and $\hat{V}_{n\underline{b}1}$ are characterized by the flow-value equations

$$\begin{aligned}
r\hat{V}_{\ell 1} &= \nu\mu_{\underline{bo}}(\hat{V}_{n\underline{s}1} - \hat{V}_{\ell 1}) \\
r\hat{V}_{n\underline{s}1} &= \overline{w}_1 + \lambda\mu_{b1}(\hat{V}_{n\underline{n}1} - \hat{V}_{n\underline{s}1}) \\
r\hat{V}_{n\underline{n}1} &= \overline{w}_1 + \underline{\kappa}(\hat{V}_{n\underline{b}1} - \hat{V}_{n\underline{n}1}) \\
r\hat{V}_{n\underline{b}1} &= \overline{w}_1 + \lambda\mu_{s1}(\hat{V}_{\ell 1} - \hat{V}_{n\underline{b}1}).
\end{aligned}$$

and the utilities $\hat{V}_{s2}$, $\hat{V}_{n2}$ are characterized by

$$\begin{aligned}
r\hat{V}_{s2} &= \delta - y + \lambda\mu_{b2}(\hat{V}_{n2} + p_2 - \hat{V}_{s2}) \\
r\hat{V}_{n2} &= \overline{\kappa}(\hat{V}_{s2} - C_2 - \hat{V}_{n2}).
\end{aligned}$$

In particular, the flow benefit $\delta - y$ is the certainty equivalent from the long position in asset 1,

which is unhedged when the arbitrageur seeks to sell asset 2. Solving these equations, we find

$$\begin{aligned} rV_{\ell 1 s 2} &= r\hat{V}_{\ell 1} + r\hat{V}_{s2} \\ &= \frac{\frac{\nu\mu_{\underline{bo}}}{r+\nu\mu_{\underline{bo}}}\left(1 - \frac{\lambda\mu_{b1}}{r+\lambda\mu_{b1}}\frac{\underline{\kappa}}{r+\underline{\kappa}}\frac{\lambda\mu_{s1}}{r+\lambda\mu_{s1}}\right)}{1 - \frac{\nu\mu_{\underline{bo}}}{r+\nu\mu_{\underline{bo}}}\frac{\lambda\mu_{b1}}{r+\lambda\mu_{b1}}\frac{\underline{\kappa}}{r+\underline{\kappa}}\frac{\lambda\mu_{s1}}{r+\lambda\mu_{s1}}}\overline{w}_1 + \left[\delta - y + \frac{\lambda\mu_{b2}}{r+\overline{\kappa}+\lambda\mu_{b2}}\left[rp_2 - \delta + y + \overline{\kappa}(p_2 - C_2)\right]\right]. \end{aligned}$$

The arbitrageur is indifferent between initiating the strategy and holding no position if $V_{\ell 1 s 2}$ is equal to p_1. Using this condition, and substituting C_1 from Equations (D.16) and (D.25)

$$\frac{\frac{\nu\mu_{\underline{bo}}}{r+\nu\mu_{\underline{bo}}}\left(1 - \frac{\lambda\mu_{b1}}{r+\lambda\mu_{b1}}\frac{\underline{\kappa}}{r+\underline{\kappa}}\frac{\lambda\mu_{s1}}{r+\lambda\mu_{s1}}\right)}{1 - \frac{\nu\mu_{\underline{bo}}}{r+\nu\mu_{\underline{bo}}}\frac{\lambda\mu_{b1}}{r+\lambda\mu_{b1}}\frac{\underline{\kappa}}{r+\underline{\kappa}}\frac{\lambda\mu_{s1}}{r+\lambda\mu_{s1}}}\overline{w}_1 = rp_1 - \delta + y - \frac{\lambda\mu_{b2}}{r+\lambda\mu_{b2}}(rp_2 - \delta + y).$$

For small search frictions, this equation becomes

$$\frac{\hat{g}_{\underline{bo}}}{r + \underline{\kappa}\frac{\hat{g}_{s1}}{r+\underline{\kappa}+\hat{g}_{s1}} + \hat{g}_{\underline{bo}}}\overline{w}_1 = r(p_1 - p_2) + \frac{r\overline{x}}{\lambda\hat{m}_{b2}},$$

and is inconsistent with Equation (26) since $w_1 > \overline{w}_1$.

Equations (25) and (26) are jointly satisfied

The two equations are jointly satisfied if

$$\frac{\hat{g}_{\underline{bo}}}{r + \underline{\kappa}\frac{\hat{g}_{s1}}{r+\underline{\kappa}+\hat{g}_{s1}} + \hat{g}_{\underline{bo}}}\frac{w_1}{r} < p_1 - p_2 < \frac{w_1}{r}.$$

Substituting p_1 and p_2 from Equations (21) and (22), we can write this equation as

$$A_1\frac{w_1}{r} < \frac{B}{\lambda} + A_2\frac{w_1}{r} < \frac{w_1}{r}, \tag{E.15}$$

where

$$A_2 \equiv \frac{\hat{g}_{\underline{bo}}}{r + \overline{\kappa} + \underline{\kappa}\frac{\hat{g}_{s1}}{r+\overline{\kappa}+\underline{\kappa}+\hat{g}_{s1}} + \hat{g}_{\underline{bo}}} < A_1 \equiv \frac{\hat{g}_{\underline{bo}}}{r + \underline{\kappa}\frac{\hat{g}_{s1}}{r+\underline{\kappa}+\hat{g}_{s1}} + \hat{g}_{\underline{bo}}} < 1$$

and

$$B \equiv \frac{(\phi r + \overline{\kappa})}{(1-\phi)}\left[\frac{1}{\hat{m}_{b2}} - \frac{1}{\hat{m}_{b1}}\right]\frac{\overline{x}}{r} > 0.$$

Equation (E.15) is satisfied if

$$\frac{B}{A_1 - A_2} > \frac{\lambda w_1}{r} > \frac{B}{1 - A_2}.$$

In this inequality, n enters only through the product λw_1. Therefore, the inequality is satisfied for some interval $n \in (n_1, n_2)$. ∎

Proof of Proposition 5: See the proof of Proposition 4. ∎

Proof of Proposition 6: We first show a small lemma.

Lemma 1 *For $\chi < 1$, inequality $(1-\chi)\hat{m}_{b1} > m_b$ is equivalent to*

$$(1 - 2\chi)(\underline{F} - \chi\overline{\kappa}\hat{m}_{b1}) > \chi\overline{F}. \tag{E.16}$$

Proof: Since m_b is the unique positive solution of Equation (B.30), whose RHS is decreasing in m_b, inequality $(1-\chi)\hat{m}_{b1} > m_b$ is equivalent to

$$\begin{aligned}
& 1 > \frac{\overline{F}}{\overline{\kappa}(1-\chi)\hat{m}_{b1} + 2\overline{\kappa}S + \frac{\overline{\kappa}+\underline{\kappa}}{\underline{\kappa}}\underline{F}} + \frac{\underline{F}}{2\overline{\kappa}(1-\chi)\hat{m}_{b1} + 2\overline{\kappa}S + \frac{\overline{\kappa}+\underline{\kappa}}{\underline{\kappa}}\underline{F}} \\
\Leftrightarrow \quad & 1 > \frac{\overline{F}}{\overline{F} + \underline{F} - \chi\overline{\kappa}\hat{m}_{b1}} + \frac{\underline{F}}{\overline{F} + \underline{F} + (1-2\chi)\overline{\kappa}\hat{m}_{b1}} \\
\Leftrightarrow \quad & \frac{\underline{F} - \chi\overline{\kappa}\hat{m}_{b1}}{\overline{F} + \underline{F} - \chi\overline{\kappa}\hat{m}_{b1}} > \frac{\underline{F}}{\overline{F} + \underline{F} + (1-2\chi)\overline{\kappa}\hat{m}_{b1}},
\end{aligned}$$

where the second step follows from Equation (B.35). It is easy to check that the last inequality implies Equation (E.16). ∎

Result (i): We need to show that $\hat{m}_{b1} > m_b$ and $\hat{g}_{s1} > g_s$. Since Equation (E.16) holds for $\chi = 0$, Lemma 1 implies that $\hat{m}_{b1} > m_b$. Using Equations (B.28) and (B.33), we can write inequality $\hat{g}_{s1} > g_s$ as

$$\frac{\overline{\kappa}S + \frac{\overline{\kappa}+\underline{\kappa}}{2\underline{\kappa}}\underline{F}}{\overline{\kappa}S + \frac{\overline{\kappa}+\underline{\kappa}}{\underline{\kappa}}\underline{F}}\hat{m}_{b1} < m_b.$$

Using Lemma 1, we then need to show that

$$(1 - 2\chi)(\underline{F} - \chi\overline{\kappa}\hat{m}_{b1}) < \chi\overline{F}, \tag{E.17}$$

for

$$\chi = \frac{\frac{\overline{\kappa}+\underline{\kappa}}{2\underline{\kappa}}\underline{F}}{\overline{\kappa}S + \frac{\overline{\kappa}+\underline{\kappa}}{\underline{\kappa}}\underline{F}}.$$

Plugging for χ, we can write Equation (E.17) as

$$\overline{\kappa}S(\underline{F} - \chi\overline{\kappa}\hat{m}_{b1}) < \frac{\overline{\kappa}+\underline{\kappa}}{2\underline{\kappa}}\underline{F}\overline{F},$$

which holds because of Equation (16) and $\hat{m}_{b1} > 0$.

Result (ii): We need to show that $\hat{m}_{b2} < m_b$ and $\hat{g}_{s2} < g_s$. Using Equations (E.8) and $\hat{m}_{b2} = \hat{m}_{\overline{b}}$, we can write inequality $\hat{m}_{b2} < m_b$ as

$$\frac{\overline{F} - \overline{\kappa}S}{\overline{F} - \overline{\kappa}S + \underline{F}}\hat{m}_{b1} < m_b.$$

Using Lemma 1, we then need to show Equation (E.17) for

$$\chi = \frac{\underline{F}}{\overline{F} - \overline{\kappa}S + \underline{F}}.$$

Plugging for χ, we can write Equation (E.17) as

$$\frac{\overline{F} - \overline{\kappa}S - \underline{F}}{\overline{F} - \overline{\kappa}S + \underline{F}}\left(\overline{F} - \overline{\kappa}S + \underline{F} - \overline{\kappa}\hat{m}_{b1}\right) < \overline{F},$$

which holds because $\hat{m}_{b1} > 0$. Using Equations (B.28), (B.31), and (E.8), we can write inequality $\hat{g}_{s2} < g_s$ as

$$\frac{\overline{F} - \overline{\kappa}S}{\overline{F} - \overline{\kappa}S + \underline{F}}\frac{\overline{\kappa}S + \frac{\overline{\kappa}+\underline{\kappa}}{2\underline{\kappa}}\underline{F}}{\overline{\kappa}S}\hat{m}_{b1} > m_b.$$

Using Lemma 1, we then need to show Equation (E.16) for

$$\chi = \frac{\underline{F}}{\overline{F} - \overline{\kappa}S + \underline{F}}\left(1 - \frac{\overline{\kappa}+\underline{\kappa}}{2\underline{\kappa}}\frac{\overline{F} - \overline{\kappa}S}{\overline{\kappa}S}\right).$$

Equation (16) implies that

$$\chi < \frac{\underline{F}}{\overline{F} - \overline{\kappa}S + \underline{F}}\left(1 - \frac{\overline{\kappa}+\underline{\kappa}}{2\underline{\kappa}}\right) < \frac{\underline{F}}{2(\overline{F} - \overline{\kappa}S + \underline{F})} \equiv \hat{\chi}.$$

Because $\hat{\chi}, \hat{m}_{b1} > 0$, Equation (E.16) holds for χ if it holds for $\hat{\chi}$. The latter is easy to check using Equation (16).

Result (iii): Equations (20), (24), and $\hat{g}_{s1} > g_s$, imply that the short-selling surplus Σ_i in the symmetric equilibrium is smaller than Σ_1 in the asymmetric equilibrium. Since, in addition, $\hat{g}_{\underline{b}o} > g_{\underline{b}o}$ (from Equations (B.27) and (B.32)), Equations (19) and (23) imply that the lending fee w_i in the symmetric equilibrium is smaller than w_1 in the asymmetric equilibrium.

Result (iv): For $\phi = 0$, the result follows from Equations (18), (21), $\hat{m}_{b1} > m_b > \hat{m}_{b2}$, $\hat{g}_{\underline{b}o} > g_{\underline{b}o}$, and the fact that the short-selling surplus Σ_i in the symmetric equilibrium is smaller than Σ_1 in the asymmetric equilibrium. An example where the prices of both assets are higher in the asymmetric equilibrium is $S = 0.5$, $\overline{F} = 3$, $\underline{F} = 5.7$, $\overline{\kappa} = 1$, $\underline{\kappa} = 3$, $\phi = \theta = 0.5$, $r = 4\%$, $\delta = 1$, $\overline{x} = 0.4$, $\underline{x} = 1.6$, $y = 0.5$, and any n.

Result (v): Social welfare is equal to the PV of the flow benefits derived by all agents. By stationarity, this is equivalent to the flow benefits derived at a given point in time. Because lending fees are a transfer, they cancel, and we only need to consider the certainty equivalents associated to the long and short positions. Summing over agents, we find

$$\sum_{i=1}^{2} \left[(\mu_{\overline{\ell}i} + \mu_{\overline{n}i}) C(\overline{\rho}, 1) + \mu_{\overline{s}i} C(0,1) + \mu_{\underline{n}i} C(\underline{\rho}, -1) + \mu_{\underline{b}i} C(0,-1) \right], \tag{E.18}$$

because long positions are held by high-valuation agents $\overline{\ell}i$ and $\overline{n}i$, and average-valuation agents $\overline{s}i$, while short positions are held by low-valuation agents $\underline{n}i$, and average-valuation agents $\underline{b}i$. Using Equations (9) and (10) to substitute for $\mu_{\overline{\ell}i}$ and $\mu_{\overline{n}i}$, and replacing the certainty equivalents by their values, we can write Equation (E.18) as

$$\sum_{i=1}^{2} \left[S(\delta + \overline{x} - y) + \mu_{\underline{n}i}(\overline{x} + \underline{x} - 2y) - \mu_{\overline{s}i}\overline{x} - \mu_{\underline{b}i}(2y - \overline{x}) \right].$$

When search frictions become small, $\mu_{\overline{s}i}$ converges to zero. To determine the limit of $\sum_{i=1}^{2} \mu_{\underline{n}i}$, we use Equation (B.8), summing over assets:

$$\sum_{i=1}^{2} \mu_{\underline{n}i} = \sum_{i=1}^{2} \frac{\nu_i \mu_{\underline{b}o} \mu_{\overline{\ell}i}}{\overline{\kappa} + \underline{\kappa}} - \sum_{i=1}^{2} \mu_{\underline{s}i}.$$

The second term in the RHS converges to zero, while the first term converges to $2g_{\underline{b}o}S/(\overline{\kappa} + \underline{\kappa})$ in the symmetric case, and $\hat{g}_{\underline{b}o}S/(\overline{\kappa} + \underline{\kappa})$ in the asymmetric case. Equations (B.27) and (B.32)

imply that in both cases the limit is $\underline{F}/\underline{\kappa}$. Therefore, the welfare comparison hinges on $\sum_{i=1}^{2} \mu_{\underline{b}i}$. Equation (B.12) implies that this converges to $\underline{F}/(\overline{\kappa} + g_s)$ in the symmetric case, and $\underline{F}/(\overline{\kappa} + \hat{g}_{s1})$ in the asymmetric case. Since $\hat{g}_{s1} > g_s$, welfare is higher in the asymmetric case. ∎

Proof of Proposition 7: Generalizing the analysis of Section B.3, we can show that a solution for $\varepsilon = 0$ exists, and is close to that for small ε. The limiting equations are (B.23) and

$$\underline{F} = \frac{\underline{\kappa}}{\overline{\kappa} + \underline{\kappa}} \sum_{i=1}^{2} \alpha_i g_{\underline{b}o} S_i, \tag{E.19}$$

$$m_{bi} = m_{\overline{b}} + \frac{\underline{\kappa}\alpha_i g_{\underline{b}o} S_i}{(\overline{\kappa} + \underline{\kappa})(\overline{\kappa} + g_{si})}, \tag{E.20}$$

$$g_{si} = \frac{\overline{\kappa} S_i}{m_{bi}} + \frac{\alpha_i g_{\underline{b}o} S_i}{m_{bi}}, \tag{E.21}$$

where the supply S_i now depends on i.

Result (i): We proceed by contradiction, assuming that for a given $S_1 - S_2 > 0$ there exists an equilibrium where $\nu_1 = \nu_2 = \nu$, even when search frictions converge to zero. Since the parameters a_i, b_i, c_i, and g_i in Equation (D.24) converge to finite limits, while f_i converges to ∞, Σ_i must converge to zero, and $f_i \Sigma_i$ to a finite limit. But then Equation (D.24) implies that the limits of c_1 and c_2 must be the same. This, in turn, implies that $g_{s1} = g_{s2} \equiv g_s$, which from Equations (E.20) and (E.21) means that

$$\frac{\overline{\kappa} S_i + g_{\underline{b}o} S_i}{m_{\overline{b}} + \frac{\underline{\kappa}\alpha_i g_{\underline{b}o} S_i}{(\overline{\kappa}+\underline{\kappa})(\overline{\kappa}+g_s)}}$$

is independent of i, a contradiction when asset supplies differ.

Result (ii): An equilibrium where $\nu_1 = \nu$ and $\nu_2 = 0$ can exist if $\Sigma_1 > 0$ and $\Sigma_2 < 0$. Condition $\Sigma_1 > 0$ can be ensured by Equation (17). For small search frictions, condition $\Sigma_2 < 0$ is equivalent to $\hat{g}_{s1} > \hat{g}_{s2}$, as shown in the proof of Proposition 4. Using Equations (E.19) and (E.21), we can write condition $\hat{g}_{s1} > \hat{g}_{s2}$ as

$$\frac{\overline{\kappa} S_1 + \frac{\overline{\kappa}+\underline{\kappa}}{\underline{\kappa}} \underline{F}}{\hat{m}_{b1}} > \frac{\overline{\kappa} S_2}{\hat{m}_{\overline{b}}}. \tag{E.22}$$

When asset supplies differ, Equation (B.35) generalizes to

$$\hat{m}_{b1} = \frac{\overline{F}}{\overline{\kappa}} - \sum_{i=1}^{2} S_i - \frac{\underline{F}}{\underline{\kappa}}, \tag{E.23}$$

and Equation (E.8) to

$$\hat{m}_{\overline{b}} = \frac{\overline{F} - \overline{\kappa} S_2}{\overline{F} - \overline{\kappa} S_2 + \underline{F}} \hat{m}_{b1}. \tag{E.24}$$

Using Equation (E.24), we can write Equation (E.22) as

$$\left[\overline{\kappa}(S_1 - S_2) + \frac{\overline{\kappa} + \underline{\kappa}}{\underline{\kappa}} \underline{F} \right] (\overline{F} - \overline{\kappa} S_2) > \overline{\kappa} S_2 \underline{F}. \tag{E.25}$$

This equation holds for all values of $S_1 \geq S_2$ because Equation (4) implies that $\overline{F} - \overline{\kappa} S_2 > \overline{\kappa} S_1 \geq \overline{\kappa} S_2$.

Result (iii): The existence condition is now (E.25), but with S_1 and S_2 reversed. It does not hold, for example, when

$$\overline{\kappa}(S_2 - S_1) + \frac{\overline{\kappa} + \underline{\kappa}}{\underline{\kappa}} \underline{F} < 0 \Leftrightarrow S_1 > S_2 + \frac{\overline{\kappa} + \underline{\kappa}}{\overline{\kappa}\underline{\kappa}} \underline{F}.$$

Result (iv): When short-selling is concentrated on asset 1, social welfare is determined by $\mu_{\underline{b}1}$, which converges to $\underline{F}/(\overline{\kappa} + \hat{g}_{s1})$ from the proof of Proposition 6. Equations (E.19), (E.21), and (E.23) imply that

$$\hat{g}_{s1} = \frac{\overline{\kappa} S_1 + \frac{\overline{\kappa} + \underline{\kappa}}{\underline{\kappa}} \underline{F}}{\frac{\overline{F}}{\overline{\kappa}} - \sum_{i=1}^{2} S_i - \frac{\underline{F}}{\underline{\kappa}}}. \tag{E.26}$$

Conversely, when short-selling is concentrated on asset 2, social welfare is determined by $\mu_{\underline{b}2}$, which converges to $\underline{F}/(\overline{\kappa} + \hat{g}_{s2})$. Moreover, $\hat{g}_{s2}$ is determined by Equation (E.26), but with S_1 and S_2 reversed. Since $S_1 > S_2$, social welfare is higher when short-selling is concentrated on asset 1. ∎

Proof of Proposition 8: With one asset in supply $2S$, the limiting equations of Section B.3

become

$$\begin{aligned}
\overline{F} &= \overline{\kappa} m_{\overline{b}} + m_{\overline{b}} g_s, \\
\underline{F} &= \frac{2\underline{\kappa}}{\overline{\kappa} + \underline{\kappa}} g_{\underline{bo}} S, \\
m_b &= m_{\overline{b}} + \frac{2\underline{\kappa} g_{\underline{bo}} S}{(\overline{\kappa} + \underline{\kappa})(\overline{\kappa} + g_s)}, \\
g_s &= \frac{2\overline{\kappa} S}{m_b} + \frac{2\gamma_{\underline{bo}} S}{m_b}.
\end{aligned}$$

Using $\tilde{m}$ and $\tilde{g}$ to denote their solution, we find

$$\begin{aligned}
\tilde{m}_b &= \frac{\overline{F}}{\overline{\kappa}} - 2S - \frac{\underline{F}}{\underline{\kappa}} = \hat{m}_{b1} > m_b > \hat{m}_{b2}, \\
\tilde{g}_s &= \frac{2\overline{\kappa} S + \frac{\overline{\kappa}+\underline{\kappa}}{\overline{\kappa}} \underline{F}}{\tilde{m}_b} > \hat{g}_{s1} > g_s > \hat{g}_{s2}, \\
\tilde{g}_{\underline{bo}} &= \frac{(\overline{\kappa} + \underline{\kappa})\underline{F}}{2\underline{\kappa} S} = g_{\underline{bo}} = \frac{\hat{g}_{\underline{bo}}}{2}
\end{aligned}$$

Proceeding as in the proof of Proposition 4, we can show that the asset price is asymptotically equal to

$$p = \frac{\delta + \overline{x} - y}{r} - \frac{\overline{\kappa}}{\lambda \tilde{m}_b} \frac{\overline{x}}{r} - \frac{\phi(r + \overline{\kappa} + \tilde{g}_s)}{\lambda(1-\phi)\tilde{m}_b} \frac{\overline{x}}{r} + \frac{\tilde{g}_{\underline{bo}} \theta \Sigma}{r}, \tag{E.27}$$

where

$$\Sigma = \frac{\underline{x} - \frac{r + \overline{\kappa} + \underline{\kappa} + \tilde{g}_s}{r + \overline{\kappa} + \tilde{g}_s}(2y - \overline{x})}{\nu(1-\theta)2S}.$$

Result (i): We compare Equations (18) and (E.27), noting that $\tilde{m}_b > m_b$, $\tilde{g}_s \tilde{m}_b = 2 g_s m_b$, $\tilde{g}_{\underline{bo}} = g_{\underline{bo}}$, and that the surplus Σ under integration exceeds Σ_i in the symmetric equilibrium because $\tilde{g}_s > g_s$.

Result (ii): To show that $p > p_2$, we compare Equations (22) and (E.27), noting that $\tilde{m}_b = \hat{m}_{b1} > \hat{m}_{b2}$ and $\tilde{g}_s \tilde{m}_b = \hat{g}_{s1} \hat{m}_{b1} + \hat{g}_{s2} \hat{m}_{b2}$. An example where $p > p_1$ is $S = 0.5$, $\overline{F} = 3$, $\underline{F} = 5.7$, $\overline{\kappa} = 1$,

$\underline{\kappa} = 3$, $\phi = \theta = 0.5$, $r = 4\%$, $\delta = 1$, $\overline{x} = 0.4$, $\underline{x} = 1.6$, $y = 0.5$, and $n = 0.2$. An example where $p < (p_1 + p_2)/2$ is for the same parameter values except $\phi = 0$.

Result (iii): This is because $\tilde{g}_s > \hat{g}_{s1}$. ∎

References

Admati, Anat R., and Paul Pfleiderer, A Theory of Intraday Patterns: Volume and Price Variability, *Review of Financial Studies*, 1988, *1*, 3–40.

___ , and ___ , Achilles and Methusaleh: The Paradox of Liquidity, Working Paper, Graduate School of Business, Stanford University, 1992.

Aiyagari, Rao, and Mark Gertler, Asset Returns with Transaction Costs and Uninsurable Individual Risks: A Stage III Exercise, *Journal of Monetary Economics*, 1991, *27*, 309–331.

Aiyagari, Rao S., Neil Wallace, and Randall Wright, Coexistence of money and interest-bearing securities, *Journal of Monetary Economics*, 1996, *37*, 397–419.

Amihud, Yakov, and Haim Mendelson, Asset Pricing and the Bid-Ask Spread, *Journal of Financial Economics*, 1986, *17*, 223–249.

___ , and ___ , Liquidity, Maturity, and the Yield on U.S. Treasury Securities, *Journal of Finance*, 1991, *46*, 479–486.

Barclay, Michael J., Terrence Hendershott, and Kenneth Kotz, Automation versus Intermediation: Evidence from Treasuries Going Off the Run, Working Paper, Haas School of Business, University of California at Berkeley, 2004.

Bennett, Paul, Kenneth Garbade, and John Kambhu, Enhancing the Liquidity of U.S. Treasury Securities in an Era of Surpluses, *Federal Reserve Bank of New York Economic Policy Review*, 2000, pp. 1–31.

Boudoukh, Jacob, and Robert F. Whitelaw, The Benchmark Effect in the Japanese Government Bond Market, *Journal of Fixed Income*, 1991, *2*, 52–59.

___ , and ___ , Liquidity as a Choice Variable: A Lesson from the Japanese Government Bond Market, *The Review of Financial Studies*, 1993, *6*, 265–292.

Brémaud, Pierre, *Point Processes and Queues*, New-York: Springer-Verlag, 1981.

Buraschi, Andrea, and David Menini, Liquidity Risk and Specialness, *Journal of Financial Economics*, 2002, *64*, 243–284.

Chowdhry, Bhagwan, and Vikram Nanda, Multimarket Trading and Market Liquidity, *Review of Financial Studies*, 1991, *4*, 483–511.

Constantinides, George M., Capital Market Equilibrium with Transaction Costs, *Journal of Political Economy*, 1986, *94*, 842–862.

Cornell, Bradford, and Alan C. Shapiro, The Misspricing of US Treasury Bonds: a Case Study, *Review of Financial Studies*, 1989, *2*, 297–310.

Diamond, Peter A., Aggregate Demand Management in Search Equilibrium, *Journal of Political Economy*, 1982, *90*, 881–894.

Duffie, Darrell, Special Repo Rates, *Journal of Finance*, 1996, *51*, 493–526.

____ , and Yeneng Sun, Existence of Independent Random Matching, Working Paper, Graduate School of Business, Stanford University, 2004.

____ , Nicolae Gârleanu, and Lasse H. Pedersen, Securities Lending, Shorting, and Pricing, *Journal of Financial Economics*, 2002, *66*, 307–339.

____ , ____ , and ____ , Valuation in Over-the-Counter Markets, Working Paper, Graduate School of Business, Stanford University, 2004.

____ , ____ , and ____ , Over-the-Counter Markets, forthcoming, *Econometrica*, 2005.

Dupont, Dominique, and Brian Sack, The Treasury Securities Market: Overview and Recent Developments, *Federal Reserve Bulletin*, 1999, *December*, 785–806.

Economides, Nicholas, and Aloysius Siow, The Division of Markets is Limited by the Extent of Liquidity, *American Economic Review*, 1988, *78*, 108–121.

Ellison, Glenn, and Drew Fudenberg, Knife Edge of Plateau: When do Markets Tip, *Quaterly Journal of Economics*, 2003, *118*, 1249–1278.

Fisher, Mark, Special Repo Rates: An Introduction, *Federal Reserve Bank of Atlanta Economic Review*, 2002, *Second Quarter*, 27–43.

Fleming, Michael J., The Round-the-Clock Market for U.S. Treasury Securities, *Federal Reserve Bank of New York Economic Policy Review*, 1997, pp. 9–32.

____ , Are Larger Treasury Issues More Liquid? Evidence from Bill Reopenings, *Journal of Money, Credit, and Banking*, 2002, *3*, 707–35.

____ , Measuring Treasury Market Liquidity, *Federal Reserve Bank of New York Economic Policy Review*, 2003, pp. 83–107.

——, and Kenneth D. Garbade, When the Back-Office Moved to the Front Burner: Settlement Fails in the Treasury Market after 9/11, *Federal Reserve Bank of New-York Economic Policy Review*, 2002, *November*, 35–57.

Goldreich, David, Bernd Hanke, and Purnendu Nath, The Price of Future Liquidity: Time-Varying Liquidity in the U.S. Treasury Market, Working Paper, Institute of Finance and Accounting, London Business School, 2002.

Graveline, Jeremy J., and Matthey R. McBrady, Who Makes the On-The-Run Treasuries Special?, Working Paper, Graduate School of Business, Stanford University, 2004.

Heaton, John, and Deborah J. Lucas, Evaluating the Effects of Incomplete Markets on Risk Sharing and Asset Pricing, *Journal of Political Economy*, 1996, *104*, 443–487.

Huang, Ming, Liquidity Shocks and Equilibrium Liquidity Premia, *Journal of Economic Theory*, 2003, *109*, 104–129.

Ibbotson, *Stock, Bonds, Bills, and Inflation Statistical Yearbook*, Chicago: Ibbotson Associate, 2004.

Jordan, Bradford D., and Susan D. Jordan, Special Repo Rates: An Empirical Analysis, *Journal of Finance*, 1997, *52*, 2051–2072.

Karatzas, Ioannis, and Steven E. Shreve, *Brownian Motion and Stochastic Calculus*, New York: Springer-Verlag, 1991.

Kiyotaki, Nobuhiro, and Randall Wright, On Money as a Medium of Exchange, *Journal of Political Economy*, 1989, *97*, 927–954.

Krishnamurthy, Arvind, The Bond/Old-Bond Spread, *Journal of Financial Economics*, 2002, *66*, 463–506.

Lo, Andrew W., Harry Mamaysky, and Jiang Wang, Asset Prices and Trading Volume under Fixed Transactions Costs, *Journal of Political Economy*, 2004, *112*, 1054–1090.

Mason, R., The 10-year Bond Markets, Credit Suisse First Boston, CSFB Research, 1987.

Moulton, Pamela C., Relative Repo Specialness in U.S. Treasuries, *Journal of Fixed Income*, 2004, *14*, 40–49.

Pagano, Marco, Endogenous Market Thinnes and Stock Price Volatility, *Review of Economic Studies*, 1989, *269-287.*

Protter, Philip, *Stochastic Integration and Differential Equations*, New York: Springer-Verlag, 1990.

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

Strebulaev, Ilya, Liquidity and Asset Pricing: Evidence from the U.S. Treasury Securities Market, Working Paper, Graduate School of Business, Stanford University, 2002.

Sundaresan, Suresh, *Fixed Income Markets and Their Derivatives*, South-Western Publishing Company, 2002.

Taylor, Angus E., and Robert W. Mann, *Advanced Calculus*, New-York: Wiley, John and Sons, 1983.

Trejos, Alberto, and Randall Wright, Search, Bargaining, Money, and Prices, *Journal of Political Economy*, 1995, *103* (1), 118–141.

Vayanos, Dimitri, Transaction Costs and Asset Prices: A Dynamic Equilibrium Model, *Review of Financial Studies*, 1998, *11*, 1–58.

____ , and Jean-Luc Vila, Equilibrium Interest Rate and Liquidity Premium with Transaction Costs, *Economic Theory*, 1999, *13*, 509–539.

____ , and Tan Wang, Search and Endogenous Concentration of Liquidity in Asset Markets, Working Paper, London School of Economics, 2004.

Wallace, Neil, A Model of the Liquidity Yield Structure Based on Asset Indivisibility, *Journal of Monetary Economics*, 2000, *45*, 55–68.

Wang, Neng, A Simple Model for Friedman's Conjecture on Consumption, Working Paper, Graduate School of Business, Columbia University, 2004.

Warga, Arthur, Bond Returns, Liquidity, and Missing Data, *Journal of Financial and Quantitative Analysis*, 1992, *27*, 605–617.

Weill, Pierre-Olivier, Liquidity Premia in Dynamic Bargaining Markets, Working Paper, Finance Department, NYU Stern School of Business, 2004.