Strategic Bargaining and Portfolio Choice in Intermediated Markets

Jessica Shi Li†

This Version: August 28, 2023

Many assets are traded in decentralized markets intermediated by dealers. In these markets, search frictions lead to trading illiquidity. Moreover, terms of trade are negotiated between investors and dealers pursuant to strategic bargaining. Investors' intrinsic types affect both their outside options and their bargaining powers. This paper proposes a search-based theory with strategic bargaining to study investors' dynamic portfolio choice and equilibrium asset prices in intermediated markets. The model rationalizes well-documented empirical patterns, and generates additional predictions novel to the literature. A key prediction is that the relationship between asset prices and liquidity is non-monotonic. In the cross-section, the price-liquidity relationship is positive for sufficiently liquid assets but negative for highly illiquid assets. The average price-liquidity relation turns negative during severe crises. The model also predicts that transaction costs are higher for investor-sell trades than for investor-buy trades. The model predictions are supported by empirical evidence using corporate bond data.

Keywords: Intermediated Assets, OTC, Portfolio Choice, Liquidity, Bargaining

JEL Codes: G01, G11, G12, G23

---

†University of Chicago Booth School of Business, email: jessicali@chicagobooth.edu. I thank Philip Bond, Jaroslav Borovicka, Doug Diamond, Maryam Farboodi, Lars Hansen, Zhiguo He, John Heaton, Doron Levit, Dan Luo, Stefan Nagel, Quentin Vandeweyer, Pierre-Olivier Weill, Randy Wright, workshop participants at the University of Chicago, and conference participants at the Summer Workshop on Money, Banking, Payments, and Finance, the Finance Theory Group Summer School and the Yiran Fan Memorial Conference for helpful comments.

1 Introduction

Many assets, including most financial assets, are traded over-the-counter (OTC) in intermediated markets. In these markets, dealers serve as intermediaries, while investors trade with dealers bilaterally. Decentralized investor-dealer trading exhibits inherent search and matching frictions, causing significant market illiquidity.[1] Importantly, the investor-dealer relationship is strategic in nature. If there is gain from trade, the investor and the dealer engage in strategic bargaining to determine the terms of trade and divide the joint trade surplus between them. The investor's intrinsic type thus affects not only their outside option (that is, their alternatives to immediate trade) but also their bargaining power (that is, their ability to capture the joint surplus). For example, an investor with imminent liquidity needs is more concerned with meeting their constraints in the short run, and the alternative to immediate trade is less attractive. Furthermore, bargaining delays can be particularly costly to such investor, and they are more willing to accept whatever terms of trade are offered to them and unable to capture much of the surplus.

Given the above features of intermediated asset markets, how does asset liquidity, defined as the ease of trading an asset, affect investors' dynamic portfolio choice and equilibrium asset prices? Specifically, how do investors allocate capital across assets with differential liquidity, and what is the relationship between asset liquidity and asset prices?

In this paper, I answer these questions by developing a highly tractable search-theoretic model with strategic bargaining. In the model, a continuum of mean-variance investors allocate capital across N assets traded in semi-centralized markets characterized by random search between investors and dealers.[2] The assets differ in their trading liquidity, due to heterogeneous search and matching frictions. Investors experience idiosyncratic shocks to their patience levels (or discount rates), giving rise to portfolio rebalancing and trading motives. The paper contributes to the theoretical literature on OTC markets in two aspects. First, I allow investors to simultaneously

[1] The literature on OTC markets primarily focuses on the search and matching frictions, which give rise to illiquidity. For example, Duffie et al. (2005, 2007), Vayanos and Weill (2008), Garleanu (2009), Lagos and Rocheteau (2009) are among some of the seminal papers.

[2] The market structure is similar to that of Lagos and Rocheteau (2009).

hold a portfolio of heterogeneous assets with unrestricted holdings. This facilitates the study of investors' portfolio choice problem in intermediated search markets. Tractability stems from the semi-centralized market structure and the assumption that assets have i.i.d. cash flows. Second, as a result of strategic bargaining, modeled as Rubinstein-style sequential bargaining, investors' patience types influence both their continuation values and their bargaining powers. The type-dependency of investors' bargaining powers marks a key departure from the existing literature.

Both asset liquidity and strategic bargaining affect investors' dynamic portfolio choice. Illiquidity exposes investors to the risk of holding imbalances, since trading is subject to delays when rebalancing needs arise. To mitigate this risk, investors attenuate their asset demand towards the average and hold less extreme positions. That is, compared to the benchmark with frictionless trading, patient investors decrease their demand while impatient investors increase their demand for the illiquid asset. Moreover and due to strategic bargaining, for a given asset, patient investors attenuate their demand more than impatient investors, leading to a demand wedge. This is because investors' bargaining powers weaken when they experience shocks and become impatient. Knowing that they will receive bad terms when selling the asset upon shocks, patient investors internalize this by demanding less of the asset to begin with.

Consequently, a key result of the model is that the relationship between asset prices and asset liquidity in intermediated markets is non-monotonic. The intuition is straightforward and stems from investors' portfolio choice. When an asset is perfectly illiquid, it would take an investor infinite amount of time to meet a dealer. Trading and thus bargaining never takes place in this limit. Because the demand wedge results from strategic bargaining, and bargaining is irrelevant in this limit, demand wedge does not arise. Demand attenuations by patient and impatient investors net out in aggregate, and asset price converges to that under the frictionless trading benchmark. If the asset's liquidity increases slightly, the effect of bargaining becomes pertinent and demand wedge emerges, causing a decline in price. However, as asset liquidity increases further, demand attenuations by both patient and impatient investors decrease, and the demand wedge gradually shrinks. Once the asset becomes sufficiently liquid, the demand wedge is small enough that the asset's price starts to rise. In the limit where the asset becomes perfectly liquid, demand attenuations and thus the demand wedge converge to zero, and the price again converges to that under the benchmark.

A novel prediction of the model is thus that the price-liquidity relationship is positive among sufficiently liquid assets but negative among highly illiquid assets. More precisely, there exists a liquidity threshold, such that among assets above the liquidity threshold, more liquid assets trade at higher prices than less liquid assets. However, among assets below the liquidity threshold, more liquid assets trade at lower prices instead. Moreover, the model suggests that the liquidity threshold increases in investors' shock intensity and shock magnitude, large values of which are often associated with market stress when investors' liquidity needs are high. Thus, the market average price-liquidity relationship may turn negative during severe crises. These results contrast with the existing literature (both theoretical and empirical) which suggests that more liquid assets should command at least weakly higher prices than less liquid assets.

Another novel prediction of the model concerns transaction costs, which have long been a main focus of the OTC markets literature and the microstructure literature. The model predicts that transaction costs are asymmetric between investor-buy and investor-sell trades. Specifically, transaction costs are higher for investors who sell to dealers than for investors who buy from dealers. The result is also intuitive. Compared to buying investors, selling investors are more impatient and have weaker bargaining powers, leading to higher transaction costs incurred by them when trading with dealers.

I test the model predictions using U.S. corporate bond data from January 2005 to December 2021. Motivated by the model, I use transaction costs to measure bond liquidity. To test the non-monotonic relationship between price and liquidity, I divide the sample into sub-samples containing relatively liquid and illiquid bonds based on common liquidity proxies. To address endogeneity concerns of regressing credit spreads on transaction costs, I exploit the institutional feature that newly-issued bonds tend to be significantly more liquid than older bonds of the same issuer, and instrument transaction costs with an instrumental variable constructed based on whether a bond is newly-issued. I show that while more liquid bonds (or bonds with lower transaction costs) trade at lower credit spreads in sub-samples containing relatively liquid bonds, the reverse is true for sub-samples containing highly illiquid bonds. In the time-series, I plot the average credit spread difference between newly-issued bonds and matched old bonds. I show that the average spread differential is generally positive, but noticeably negative from Q4 2008 to Q2 2009 following the Lehman Brothers collapse and in the depth of the Great Financial Crisis. Finally, I examine the transaction cost differential between investor-buy and investor-sell trades, and find that selling investors incur on average 3-6 bps higher transaction costs than buying investors.

1.1 Related Literature

This paper is related to the literature on OTC markets. Early models in the literature (e.g., Duffie et al. (2005, 2007), Vayanos and Weill (2008), Weill (2008)) assume that investors are risk-neutral and hold zero or one unit of an asset. Lagos and Rocheteau (2009) relax investors' holding restrictions, while maintaining tractability by modeling the trading market as semi-centralized. That is, investor-dealer trading is bilateral, but dealers have continuous access to a competitive inter-dealer market. I follow their approach. The paper thus falls under the body of literature with risk-averse agents and unrestricted holdings (e.g., Garleanu (2009), Afonso and Lagos (2015), Üslü (2019)). The paper makes two theoretical contributions.

First, this paper explicitly models investors' portfolio choice problem. In the model, investors can simultaneously hold a portfolio of intermediated assets with arbitrary holdings. This contrasts with existing multi-asset models of OTC markets (e.g., Vayanos and Wang (2007), Vayanos and Weill (2008), Weill (2008), Milbradt (2017), An (2022), Sambalaibat (2022)), where investors can only search for one asset at a time. A recent paper by Üslü and Velioğlu (2021) also allows investors to hold portfolios of OTC assets. In their model, trading is fully decentralized, and the authors focus instead on liquidity outcomes (e.g., price dispersion, trading volume and price impact), rather than the relationship between liquidity and asset prices, which is the focus of this paper.

Second, motivated by the asset pricing literature, I model investors' idiosyncratic shocks as discount rate shocks, which in turn affect the equilibrium outcomes of strategic bargaining between investors and dealers. Similar to Duffie et al. (2007) and Farboodi et al. (2019), I model the investor-dealer bargaining problem as a sequential bargaining game with alternating offers (e.g., Rubinstein (1982), Rubinstein and Wolinsky (1985)). However, the solutions to the strategic bargaining problem in my model are no longer the same as those under the axiomatic approach to bargaining, as investors with higher discount rates have not only lower continuation values but also weaker bargaining powers. In other words, investors' intrinsic types affect both their bargaining powers and their outside options. The type-dependent bargaining powers, which result from strategic bargaining, mark a key departure from the literature.

This paper also contributes to a large literature in asset pricing that examines the effect of illiquidity on asset prices, including earlier theoretical work such as Amihud and Mendelson (1986), Constantinides (1986), Aiyagari and Gertler (1991), Heaton and Lucas (1996), Vayanos (1998), and more recent empirical studies such as Pastor and Stambaugh (2003), Acharya and Pedersen (2005). The paper is particularly relevant to the growing literature on liquidity and asset prices in fixed income markets, for example, Longstaff et al. (2005), Chen et al. (2007), Lin et al. (2011), Acharya et al. (2013) among others for the corporate bond market, and Amihud and Mendelson (1991), Pasquariello and Vega (2009), Favero et al. (2010) among others for the Treasuries market. While the common belief in the existing literature is that liquid assets should trade at higher prices than illiquid assets, a few recent empirical papers find that liquidity premium appears to be negative during severe market stress (e.g., Boudoukh et al. (2019), Choi et al. (2019)). A concurrent paper by Choi et al. (2022) documents that the price-liquidity relationship in the corporate bond market is positive during normal periods but negative during crises. The authors construct a model based on Vayanos and Wang (2007) to explain the phenomenon. In their model, trading is concentrated in the more liquid asset, and when sellers are the marginal investors, they drive down the price of the more liquid asset. The mechanism of my model is thus very different. Importantly, the non-monotonic relationship between asset liquidity and prices in my model is a cross-sectional statement, and the price-liquidity relationship is negative among assets that are sufficiently illiquid. The negative relationship between asset liquidity and prices simply becomes more prevalent during times of stress.

The rest of this paper is organized as follows. Section 2 describes the model setup. Section 3 derives the stationary equilibrium in closed form, and establishes existence and uniqueness. Section 4 discusses the main equilibrium properties and presents key model predictions. Section 5 tests the model predictions empirically using corporate bond transaction data. Section 6 concludes.

2 Model Environment

Time is continuous and infinite, with $t \in [0, \infty)$. I fix a probability space $(\Omega, \mathcal{F}, \Pr)$ and a filtration $\{\mathcal{F}_t : t \geq 0\}$ of sub-σ-algebras satisfying the usual conditions. The filtration represents the resolution of commonly available information over time. The economy is populated by a continuum of agents, who produce and consume a perishable numéraire good. Agents enjoy consuming the numéraire good, with marginal utility of consumption normalized to one.

2.1 Assets

There are N long-lived risky assets, indexed by $i \in \{1, 2, \dots, N\}$. Each asset is in fixed supply $s > 0$. From time t to $t + dt$, asset i produces dD_t^i units of consumption in the form of dividends. For tractability, I assume that the assets have i.i.d. dividend flows.[3] Asset i's cumulative dividend flow D_t^i follows the Brownian motion

$$dD_t^i = \bar{D}dt + \sigma dZ_t^i \tag{2.1}$$

where $\bar{D} > 0$ and $\sigma > 0$ are constants, and Z_t^i is a standard Brownian motion with respect to the given probability space and filtration. For any two assets $i \neq i'$, the standard Brownian motions Z_t^i and $Z_t^{i'}$ are independent.

[3]The assumption of i.i.d. dividend flows provides tractability. The mechanism remains unchanged if the assets have correlated dividend processes.

The assets are traded over-the-counter in semi-centralized markets intermediated by dealers. I normalize the measure of dealers to one. The market structure resembles that of Lagos and Rocheteau (2009). Investors must trade with dealers, and investor-dealer trading is fully bilateral. Assets are traded individually by dealers, in the sense that each dealer can trade only one of the N assets with an investor in a given trading session. By contrast, dealers can continuously trade in a competitive inter-dealer market. Investors thus have indirect access to the inter-dealer market through dealers. Dealers discount time at a constant rate $r > 0$. Dealers do not hold inventories, and their instantaneous utility is their numéraire consumption.[4]

Due to imperfect search and matching typical in over-the-counter (OTC) markets, investor trading is not instantaneous but subject to delays. Following the random search framework, I assume that investors wishing to trade asset i randomly contact dealers and are met with a dealer at independent Poisson arrival times with intensity $\lambda^i > 0$. Hence, the average trading delay for asset i is $1/\lambda^i$. The meeting intensity λ^i, an asset-level characteristic, thus captures the asset's liquidity defined as the ease of trading it and related to trading frictions. An asset with higher meeting intensity is more liquid, as trading it takes less time thanks to more effective search and matching. Without loss of generality, I assume that $\lambda^1 > \lambda^2 > \cdots > \lambda^N$.

2.2 Investors

There is a continuum of investors in the economy. Without loss of generality, the total measure of investors is assumed to be one. Investors make portfolio allocation decisions among the N assets. An investor holding $\mathbf{x} \equiv (x^1, x^2, \ldots, x^N)$ units of the assets derives mean-variance flow benefit of

$$u(\mathbf{x}) = \sum_{i=1}^{N} \bar{D}x^i - \frac{\gamma}{2}\sum_{i=1}^{N} \sigma^2 (x^i)^2 \tag{2.2}$$

[4]The restriction on dealer inventories can be relaxed by assuming that dealers derive utility (e.g., a mean-variance benefit) from holding assets, such that dealers' optimal asset holdings are non-zero. Relaxing the restriction complicates the exposition, without affecting the economics. I thus keep the restriction that dealers hold no inventories, since it is immaterial for steady-state analysis and imparts all the key intuition.

$\bar{D}x^i$ is the instantaneous expected dividend payment from the investor's asset i holding, and $\sigma^2(x^i)^2$ is the instantaneous variance of the dividend payment. Because the assets' dividend flows are i.i.d., the two sums on the right-hand side of (2.2) are the instantaneous mean and variance of the portfolio cash flows respectively. The coefficient γ captures investors' risk aversion towards fundamental risks. A large γ means that investors are more averse to cash flow uncertainty. Investors are risk-neutral towards jump risks such as transition in types and arrival of trading opportunities.[5] The instantaneous utility function of an investor has the quasi-linear form $u(\mathbf{x})+c$, where $c \in \mathbb{R}$ is the investor's numéraire consumption. In the model, investors could be interpreted as mean-variance optimizers of short-term trading profits. In practice, most institutional investors are subject to frequent mark-to-market for performance measurement and compliance purposes.

Investors experience idiosyncratic patience shocks, arising from imminent liquidity needs. For example, an investor facing significant fund outflows must raise cash quickly, in order to meet the looming redemption deadlines. I assume that an investor is either in a normal state or a shock state. An investor in the normal state is patient and discounts time at the rate r. An investor in the shock state becomes impatient and discounts time at a higher rate $r + \epsilon$, where $\epsilon > 0$. Thus, impatient investors assign greater value to immediate consumption. Such investors are more likely to consume, than to hold onto the asset for future consumption. Consequently, impatient investors have lower asset demand than patient investors.

I use $\xi \in \{h, l\}$ to denote an investor's patience type. A high type (or type h) investor is in the normal state and is patient. A low type (or type l) investor is in the shock state and is impatient. Investors transition, randomly and pairwise-independently, from type h to type l with intensity $\zeta_{hl} > 0$, and from type l to type h with intensity $\zeta_{lh} > 0$. Switching of patience types gives rise to trading motives by investors. Importantly, when an investor's type changes, they wish to adjust their holdings in all portfolio assets. Let π denote the proportion of investors that are impatient in the steady state. It is immediate that the steady-state proportions of type l and

[5] The specification of mean-variance investors follows Üslü (2019), and can be derived from the specification where investors maximize CARA utility with respect to their intertemporal consumption, up to a suitable first-order approximation (e.g., Duffie et al. (2007), Vayanos and Weill (2008), Garleanu (2009), Üslü (2019). The approximation retains risk aversion towards fundamental risks, and linearizes preferences over jumps. Garleanu (2009) provides numerical examples and demonstrates the accuracy of the approximation under reasonable parameter assumptions.

type h investors are respectively given by

$$\pi = \frac{\zeta_{hl}}{\zeta_{hl} + \zeta_{lh}} \quad \text{and} \quad 1 - \pi = \frac{\zeta_{lh}}{\zeta_{hl} + \zeta_{lh}} \tag{2.3}$$

2.3 Strategic Bargaining

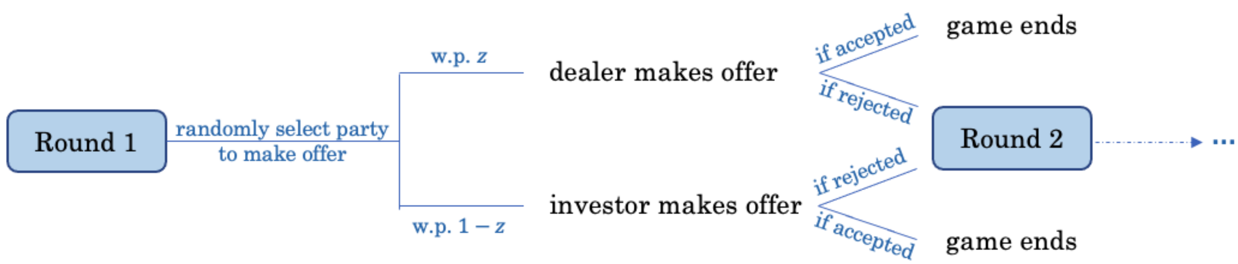
When an investor and a dealer meet in a trading session for asset $i \in \{1, 2, \dots, N\}$, they enter into a Rubinstein-style bargaining game that occurs in virtual time.[6] The terms of trade, including the trade price and the trade quantity, are determined by bargaining. The bargaining game is sequential with alternating offers, à la Rubinstein and Wolinsky (1985). The timing of the bargaining game is illustrated in Figure 1. At the beginning of a bargaining round, one party is selected at random to suggest a partition of the joint trade surplus. Specifically, the dealer is selected to make an offer with probability $z \in (0, 1)$, and the investor is selected with probability $1 - z$.[7] The other party immediately accepts or rejects the offer. The bargaining game ends if the offer is accepted. Otherwise, the game continues with the initiation of a new round, again with one party randomly selected (dealer with probability z and investor with probability $1 - z$) to make an offer. The game begins with round one, and continues until an offer is accepted and the game ends.

During a trading session between an investor and a dealer, the investor and the dealer bargain over the division of the joint trade surplus. The actual division rule is an equilibrium outcome of the strategic bargaining game described above. As will become clear later, a key feature of the model is that an investor's patience type affects not only their outside option, but also their bargaining power defined as the fraction of the joint trade surplus the investor can capture.

[6] The assumption that bargaining occurs in virtual time greatly simplifies the model, while keeping the mechanism unchanged. Given the assumption, the investor does not switch types during bargaining and there is no need to keep track of changing discount rates. The assumption is also realistic, as bargaining in practice takes little time compared to trading delays and investors' time horizons.

[7] Dealers' offer probability z is assumed to be exogenous. It depends on dealers' general market powers and bargaining skills, which are outside the model.

Figure 1: Bargaining Game



Note: The figure illustrates the timing of the sequential bargaining game between an investor and a dealer. In each round of the bargaining game, the dealer is randomly selected with probability z and the investor is randomly selected with probability $1 - z$ to make an offer. If the other party accepts the offer, the trade takes place and the game ends. Otherwise, the game proceeds to the next round.

Patient (type h) investors enjoy greater bargaining powers than impatient (type l) investors and extract greater surplus from trade.

3 Equilibrium Analysis

In this section, I characterize and derive the stationary equilibrium for the economy.

3.1 Value Functions

Given my focus on the stationary equilibrium, I henceforth suppress the time argument, unless otherwise specified. Let $V_\xi(\mathbf{x})$ denote the expected value of an investor with patience type $\xi \in \{h, l\}$ and asset holdings $\mathbf{x} \equiv (x^1, x^2, \ldots, x^N) \in \mathbb{R}^N$. Let $P^i_\xi(\mathbf{x})$ and $q^i_\xi(\mathbf{x})$ denote the trade price and the trade quantity of asset i when the investor meets a dealer in a trading session for the asset.

Consider the value function of a type h investor with asset holdings $\mathbf{x}$. From time t to $t + dt$, the investor receives a mean-variance benefit of $u(\mathbf{x})dt$ from their asset holdings. The investor switches from patience type h to patience type l with probability $\zeta_{hl} dt$. The investor meets a dealer for the asset i with probability $\lambda^i dt$. Thus, $V_h(\mathbf{x})$ at time t can be written as the expected value of the investor at time $t + dt$, discounted at the investor's time-t discount rate r. That is,

$$\begin{aligned} V_h(\mathbf{x}) = (1 - rdt)\Big\{ & u(\mathbf{x})dt + \Big(1 - \zeta_{hl}dt - \sum_{i=1}^{N} \lambda^i dt\Big) V_h(\mathbf{x}) + \zeta_{hl} dt V_l(\mathbf{x}) \\ & + \sum_{i=1}^{N} \lambda^i dt \Big[V_h\big(x^1, \ldots, x^i + q_h^i(\mathbf{x}), \ldots, x^N\big) - P_h^i(\mathbf{x}) q_h^i(\mathbf{x}) \Big] \Big\} \end{aligned} \tag{3.1}$$

Rearrange (3.1), the investor's value function satisfies the following Hamilton-Jacobi-Bellman (HJB) equation

$$\begin{aligned} rV_h(\mathbf{x}) = & \sum_{i=1}^{N} \bar{D} x^i - \frac{\gamma}{2} \sum_{i=1}^{N} \sigma^2 (x^i)^2 + \zeta_{hl} \Big[V_l(\mathbf{x}) - V_h(\mathbf{x}) \Big] \\ & + \sum_{i=1}^{N} \lambda^i \Big[V_h\big(x^1, \ldots, x^i + q_h^i(\mathbf{x}), \ldots, x^N\big) - V_h(\mathbf{x}) - P_h^i(\mathbf{x}) q_h^i(\mathbf{x}) \Big] \end{aligned} \tag{3.2}$$

The first two terms on the right-hand side correspond to the mean-variance flow benefit derived by the investor. The bracketed term $V_l(\mathbf{x}) - V_h(\mathbf{x})$ represents the change in value due to switching of patience type. The bracketed term in the summand represents the change in value due to trading asset i. Analogously, the HJB equation for a type l investor with asset holdings $\mathbf{x}$ is

$$\begin{aligned} (r + \epsilon)V_l(\mathbf{x}) = & \sum_{i=1}^{N} \bar{D} x^i - \frac{\gamma}{2} \sum_{i=1}^{N} \sigma^2 (x^i)^2 + \zeta_{lh} \Big[V_h(\mathbf{x}) - V_l(\mathbf{x}) \Big] \\ & + \sum_{i=1}^{N} \lambda^i \Big[V_l\big(x^1, \ldots, x^i + q_l^i(\mathbf{x}), \ldots, x^N\big) - V_l(\mathbf{x}) - P_l^i(\mathbf{x}) q_l^i(\mathbf{x}) \Big] \end{aligned} \tag{3.3}$$

3.2 Bargaining

When a type ξ investor with asset holdings $\mathbf{x}$ meets a dealer in a trading session for asset i, the trade price $P_\xi^i(\mathbf{x})$ and the trade quantity $q_\xi^i(\mathbf{x})$ are solutions to the strategic bargaining game described in Section 2.3. Given the trade price and the trade quantity, the investor's gain from trade is $V_\xi\big(x^1, \ldots, x^i + q_\xi^i(\mathbf{x}), \ldots, x^N\big) - V_\xi(\mathbf{x}) - P_\xi^i(\mathbf{x}) q_\xi^i(\mathbf{x})$. Let $\bar{P}^i$ denote asset i's inter-dealer clearing price, the main object of interest in this paper. Then the dealer's gain from trade is $P_\xi^i(\mathbf{x}) q_\xi^i(\mathbf{x}) - \bar{P}^i q_\xi^i(\mathbf{x})$, the dealer's intermediation profit for trading asset i. Hence, the joint trade surplus between the investor and the dealer is the sum of their respective gains from trade, $V_\xi\big(x^1, \ldots, x^i + q^i_\xi(\mathbf{x}), \ldots, x^N\big) - V_\xi(\mathbf{x}) - \bar{P}^i q^i_\xi(\mathbf{x})$.

The investor and the dealer bargain over the trade price and the trade quantity to divide the join trade surplus between them. In a given bargaining round, one party is randomly selected to suggest a price and a quantity to split the joint surplus. The suggested terms of trade are such that the other party is indifferent between accepting and rejecting the offer. The solutions to the bargaining game with respect to asset i are presented in the following lemma. The detailed derivations are in Appendix A.1.

Lemma 1. *When an investor with patience type $\xi \in \{h, l\}$ and asset holdings $\mathbf{x} = (x^1, x^2, \ldots, x^N)$ meets a dealer in a trading session for asset i, the trade price $P^i_\xi(\mathbf{x})$ and the trade quantity $q^i_\xi(\mathbf{x})$ satisfy*

$$V_{\xi i}\big(x^1, \ldots, x^i + q^i_\xi(\mathbf{x}), \ldots, x^N\big) = \bar{P}^i \tag{3.4}$$

$$P^i_\xi(\mathbf{x}) q^i_\xi(\mathbf{x}) = (1 - \theta_\xi)\Big[V_\xi\big(x^1, \ldots, x^i + q^i_\xi(\mathbf{x}), \ldots, x^N\big) - V_\xi(\mathbf{x})\Big] + \theta_\xi \bar{P}^i q^i_\xi(\mathbf{x}) \tag{3.5}$$

where the subscript i of $V_{\xi i}$ in (3.4) denotes derivative of V_ξ with respect to the i-th argument, and

$$\theta_\xi = \frac{(1-z)r}{r + z\epsilon \mathbb{I}_{\{\xi = l\}}} \tag{3.6}$$

Equation (3.4) says that the trade quantity $q^i_\xi(\mathbf{x})$ equates the post-trade marginal value of the investor to the inter-dealer price of the asset. As discussed previously, when the inter-dealer market is competitive, investors can be viewed as having indirect access to the inter-dealer market through dealers. Thus, the inter-dealer price is also the marginal cost of purchasing the asset. Hence, equation (3.4) means that the marginal value from trading must equal the marginal cost. Note that the trade quantity that solves (3.4) also maximizes the joint trade surplus between the investor and the dealer. If the actual trade quantity deviated from $q^i_\xi(\mathbf{x})$, there would be Pareto improvement by setting the quantity to $q^i_\xi(\mathbf{x})$ instead.

Equation (3.5) says that the trade price $P^i_\xi(\mathbf{x})$ and the trade quantity $q^i_\xi(\mathbf{x})$ divide the joint trade surplus, in such a way that the type ξ investor receives a fraction θ_ξ of the joint surplus. To see this, substitute (3.5) into the investor's gain from trade. The investor's gain from trade can thus be written as $\theta_\xi[V_\xi\big(x^1, \ldots, x^i + q^i_\xi(\mathbf{x}), \ldots, x^N\big) - V_\xi(\mathbf{x}) - \bar{P}^i q^i_\xi(\mathbf{x})]$. That is, the investor receives a fraction θ_ξ of the joint trade surplus between the investor and the dealer. Similarly, the dealer receives a fraction $1 - \theta_\xi$ of the joint trade surplus. Thus, θ_ξ and $1 - \theta_\xi$ correspond to the investor's and the dealer's relative bargaining powers respectively.

The bargaining powers are equilibrium outcomes of the strategic bargaining problem. The investor's bargaining power θ_ξ is given by (3.6). The indicator $\mathbb{I}_{\{\xi=l\}}$ equals one if the investor is type l, and zero otherwise. The investor's bargaining power depends on their patience type ξ. From (3.6), it is immediate that $\theta_h > \theta_l$. That is, patient investors have stronger bargaining powers than impatient investors, and extract greater surplus from trade. Furthermore, the more impatient a type l investor, the worse their bargaining power is relative to patient investors, since θ_l decreases in ϵ. In the extreme case that an investor only values current consumption (i.e., infinitely impatient), the investor is willing to accept an offer with zero surplus from trade.

Equations (3.4) and (3.5) are solutions to the strategic bargaining game described earlier. They are, however, analogous to the Nash bargaining results typical in the literature on OTC markets. The difference is that under strategic bargaining, investors' bargaining powers are endogenous and depend on their types. By contrast, bargaining powers are exogenous under the axiomatic approach commonly followed in the literature. The above results thus suggest that the investor's patience type affects both their bargaining power and their outside option. Because an impatient investor faces more costly bargaining delays off-equilibrium, they are more willing to accept whatever deal is offered to them by the dealer. Hence in equilibrium, the investor receives worse terms of trade and captures less of the joint trade surplus. At the same time, the impatient investor discounts the future at a higher rate, leading to lower continuation value and diminished outside option.

3.3 Walrasian Benchmark

Before proceeding further, I present the stationary equilibrium of a benchmark economy where all assets are traded in frictionless Walrasian markets. The Walrasian benchmark serves as the basis for comparison and highlights the effects of frictions. In the benchmark economy, investors optimize while taking the Walrasian market prices $\bar{P}^{1,w}, \bar{P}^{2,w}, \ldots, \bar{P}^{N,w}$ as given. Investors have the usual demand functions of mean-variance agents (details are contained in Appendix B). The optimal asset i demand by type h and type l investors are respectively given by

$$x_h^{i,w} = \frac{\bar{D} - r\bar{P}^{i,w}}{\gamma\sigma^2} \quad \text{and} \quad x_l^{i,w} = \frac{\bar{D} - (r+\epsilon)\bar{P}^{i,w}}{\gamma\sigma^2} \tag{3.7}$$

for all $i \in \{1, 2, \ldots, N\}$. Market clearing of asset i requires that $(1-\pi)x_h^{i,w} + \pi x_l^{i,w} = s$, where π is the steady-state proportion of impatient investors as given by (2.3). Thus, the asset's market price is determined as

$$\bar{P}^{i,w} = \frac{\bar{D} - \gamma\sigma^2 s}{r + \pi\epsilon} \tag{3.8}$$

3.4 Equilibrium

By substituting (3.5) into investors' HJB equations (3.2) and (3.3), I rewrite the HJB equations of type h and type l investors respectively as

$$\begin{aligned} rV_h(\mathbf{x}) = &\sum_{i=1}^{N} \bar{D}x^i - \frac{\gamma}{2}\sum_{i=1}^{N} \sigma^2 (x^i)^2 + \zeta_{hl}\Big[V_l(\mathbf{x}) - V_h(\mathbf{x})\Big] \\ &+ \sum_{i=1}^{N} \lambda^i \theta_h \Big[V_h\big(x^1, \ldots, x^i + q_h^i(\mathbf{x}), \ldots, x^N\big) - V_h(\mathbf{x}) - \bar{P}^i q_h^i(\mathbf{x})\Big] \end{aligned} \tag{3.9}$$

and

$$
\begin{aligned}
(r+\epsilon)V_l(\mathbf{x}) = &\sum_{i=1}^{N} \bar{D}x^i - \frac{\gamma}{2}\sum_{i=1}^{N}\sigma^2(x^i)^2 + \zeta_{lh}\Big[V_h(\mathbf{x}) - V_l(\mathbf{x})\Big] \\
&+ \sum_{i=1}^{N}\lambda^i\theta_l\Big[V_l\big(x^1,\ldots,x^i+q_l^i(\mathbf{x}),\ldots,x^N\big) - V_l(\mathbf{x}) - \bar{P}^i q_l^i(\mathbf{x})\Big]
\end{aligned}
\tag{3.10}
$$

To solve for investors' expected values, I use a guess-and-verify approach. I conjecture (and verify) that $V_h(\mathbf{x})$ and $V_l(\mathbf{x})$ are quadratic in their arguments. I then substitute the conjectured value functions into (3.9) and (3.10), and differentiate with respect to x^i to get the expressions for investors' marginal values from trading asset i. The solutions for the marginal values are then obtained using the method of undetermined coefficients. As shown in Appendix A.2, the marginal values of type h and type l investors with respect to asset $i \in \{1, 2, \ldots, N\}$ satisfy

$$
V_{hi}(\mathbf{x}) - \bar{P}^i = \frac{(r+\epsilon+\zeta_{lh}+\zeta_{hl}+\lambda^i\theta_l)(\bar{D}-\gamma\sigma^2x^i - r\bar{P}^i) - \epsilon\zeta_{hl}\bar{P}^i}{(r+\zeta_{hl}+\lambda^i\theta_h)(r+\epsilon+\zeta_{lh}+\lambda^i\theta_l) - \zeta_{hl}\zeta_{lh}} \tag{3.11}
$$

$$
V_{li}(\mathbf{x}) - \bar{P}^i = \frac{(r+\zeta_{lh}+\zeta_{hl}+\lambda^i\theta_h)[\bar{D}-\gamma\sigma^2x^i - (r+\epsilon)\bar{P}^i] + \epsilon\zeta_{lh}\bar{P}^i}{(r+\zeta_{hl}+\lambda^i\theta_h)(r+\epsilon+\zeta_{lh}+\lambda^i\theta_l) - \zeta_{hl}\zeta_{lh}} \tag{3.12}
$$

Due to the i.i.d. dividend flows assumption, investors' other asset holdings do not enter the right-hand sides of (3.11) and (3.12). Since an investor trading asset i with a dealer has post-trade marginal value equal to the asset's inter-dealer price, the investor's optimal asset i holding depends only on their patience type. In other words, type h investors' optimal asset i holding x_h^i is such that the indifference condition $V_{hi}(\mathbf{x})|_{x^i=x_h^i} = \bar{P}^i$ holds, and type l investors' optimal asset i holding x_l^i is such that $V_{li}(\mathbf{x})|_{x^i=x_l^i} = \bar{P}^i$. The optimal asset i holdings of type h and type l investors can be obtained by substituting the indifference conditions into (3.11) and (3.12) evaluated at x_h^i and x_l^i respectively. I present the results in the proposition below. The detailed derivations and proofs can be found in Appendix A.2.

Proposition 2. *For all $i \in \{1, 2, \ldots, N\}$, the optimal asset i holdings by type h and type l investors are respectively*

$$x_h^i = \frac{\bar{D} - (r + \frac{\zeta_{hl}\epsilon}{r+\epsilon+\zeta_{lh}+\zeta_{hl}+\lambda^i\theta_l})\bar{P}^i}{\gamma\sigma^2} \tag{3.13}$$

$$x_l^i = \frac{\bar{D} - (r + \epsilon - \frac{\zeta_{lh}\epsilon}{r+\zeta_{lh}+\zeta_{hl}+\lambda^i\theta_h})\bar{P}^i}{\gamma\sigma^2} \tag{3.14}$$

According to Proposition 2, the optimal asset i holdings for type h and type l investors are x_h^i and x_l^i respectively, for all i. An investor has incentive to trade an asset, if their holding in that asset deviates from the desired level. Upon trading, the investor rebalances their asset holding to the optimal level. The investor's holding in the asset again becomes sub-optimal, when the investor's patience type switches. Thus, in the steady state, there are only two holding levels for each asset i, x_h^i and x_l^i. It is easy to verify that $x_h^i > x_l^i$, that is, patient investors have higher asset demand than impatient investors. Hence, in equilibrium, patient investors holding x_l^i units of asset i wish to purchase asset i, while impatient investors holding x_h^i units of asset i wish to sell the asset. Notice that so far I have assumed that when an investor trades with a dealer, the dealer knows the investor's patience type. In this model, because buying investors must be patient and selling investors must be impatient, investors' patience types are perfectly revealed by their trading behavior.

Compared to the Walrasian benchmark, investors' optimal asset holdings are less extreme when assets are illiquid. From (3.13), a type h investor's optimal asset i demand is $\frac{\zeta_{hl}\epsilon}{r+\epsilon+\zeta_{lh}+\zeta_{hl}+\lambda^i\theta_l}\frac{\bar{P}}{\gamma\sigma^2}$ lower than that in the benchmark case $x_h^{i,w}$. From (3.14), a type l investor's optimal asset i demand is $\frac{\zeta_{lh}\epsilon}{r+\zeta_{lh}+\zeta_{hl}+\lambda^i\theta_h}\frac{\bar{P}}{\gamma\sigma^2}$ higher than that in the benchmark case $x_l^{i,w}$. That is, when an asset is illiquid, investors attenuate their demand for the asset towards the average so that the desired holdings by patient and impatient investors are closer to each other. Moreover, patient investors attenuate their demand more than impatient investors. Intuitively, illiquidity exposes investors to the risk of holding imbalances, since trading is subject to delays when rebalancing needs arise. Investors thus "hedge" against this risk by holding less extreme positions. Naturally, the hedging motive decreases with asset liquidity.[8] On the other hand, asymmetric demand attenuations by patient and impatient investors result from strategic bargaining. Investors' bargaining powers weaken when they experience patience shocks and become impatient. Knowing that they will receive bad terms when trading upon shocks, patient investors demand less of the asset to begin with, in order to internalize the effect of weaker bargaining powers in the shock state.

Let $\Phi(\mathbf{x}, \xi)$ denote the steady-state joint distribution of investors' asset holdings and patience types. There are two patience types $\xi \in \{h, l\}$. As discussed above, there are only two holding levels for each asset in the steady state, namely x_h^i and x_l^i for $i \in \{1, 2, \ldots, N\}$. Hence, $(\mathbf{x}, \xi) \in \mathcal{T} = \{x_h^1, x_l^1\} \times \{x_h^2, x_l^2\} \times \cdots \times \{x_h^N, x_l^N\} \times \{h, l\}$. Let $\mu(x^1, x^2, \ldots, x^N, \xi)$ denote the steady-state measure of investors with asset holdings $(x^1, x^2, \ldots, x^N)$ and patience type ξ, where $x^i \in \{x_h^i, x_l^i\}$ and $\xi \in \{h, l\}$. Stationarity of investors' measures requires that

$$\begin{aligned} 0 = &- \zeta_{\xi\xi'} \mu(x^1, x^2, \ldots, x^N, \xi) + \zeta_{\xi'\xi} \mu(x^1, x^2, \ldots, x^N, \xi') \\ &- \sum_{i=1}^{N} \lambda^i \mu(x^1, \ldots, x^i, \ldots, x^N, \xi) \mathbb{I}_{\{x^i \neq x_\xi^i\}} \\ &+ \sum_{i=1}^{N} \lambda^i \mu(x^1, \ldots, x^{i\prime}, \ldots, x^N, \xi) \mathbb{I}_{\{x^i = x_\xi^i\}} \end{aligned} \tag{3.15}$$

where $\xi' \neq \xi \in \{h, l\}$ and $x^{i\prime} \neq x^i \in \{x_h^i, x_l^i\}$ for all $i \in \{1, 2, \ldots, N\}$. The stationarity condition says that in the steady state, the measure of investors with asset holdings $\{x^1, x^2, \ldots, x^N\}$ and patience type ξ must be constant. $\zeta_{\xi\xi'}$ is the intensity at which an investor transitions from patience type ξ to patience type ξ'. The first two terms on the right-hand side of (3.15) are thus outflow from and inflow into the investor population due to patience type switches. The third term on the right-hand side is the outflow from the investor population due to trade. $\mathbb{I}_{\{x^i \neq x_\xi^i\}}$ is an indicator that equals one if asset i holding x^i is not optimal, and zero otherwise. Investors with sub-optimal asset i holding wish to trade, and they come across a dealer at rate λ^i. Investors with optimal asset i holding do not trade the asset. Similarly, the last term on the right-hand side

[8] In the limit with $\lambda^i \to \infty$, trading delays in asset i become infinitesimally small, investors' demand attenuations (relative to the Walrasian benchmark) converge to zero, and the optimal asset holdings converge to those in the benchmark case.

is the inflow due to trade. $\mathbb{I}_{\{x^i = x^i_\xi\}}$ is an indicator equal to one if the asset i holding is optimal, and thus $x^{i'}$ is sub-optimal.

Since dealers do not hold any inventory, the entire supply of an asset is held by investors. Market clearing of asset i, where $i \in \{1, 2, \ldots, N\}$, thus requires that

$$\mathbb{E}_\Phi[x^i] = s \qquad\qquad \forall\, i \tag{3.16}$$

where $\mathbb{E}_\Phi$ denotes expectation over the steady-state joint distribution $\Phi(\mathbf{x}, \xi)$ of investors' asset holdings and patience types. The left-hand side of (3.16) is the mean asset i holding by investors. Since there is a measure one of investors, it equals the aggregate asset i holding by investors. Market clearing requires that the aggregate asset i holding must equal the asset's supply s.

Definition 1. *A stationary equilibrium is a collection of (i) inter-dealer prices $\{\bar{P}^i\}_{i\in\{1,2,\ldots,N\}}$, (ii) pricing functions $P^i : \mathcal{T} \to \mathbb{R}$, $\forall\, i \in \{1, 2, \ldots, N\}$, (iii) trade quantity functions $q^i : \mathcal{T} \to \mathbb{R}$, $\forall\, i \in \{1, 2, \ldots, N\}$, and (iv) joint distribution $\Phi(\mathbf{x}, \xi)$ of investors' asset holdings and patience types, such that*

1. *Optimality: given (i)-(iv), the expected values of type h and type l investors satisfy the HJB equations* (3.2) *and* (3.3) *respectively;*
2. *Bargaining: (ii) and (iii) are solutions to the sequential bargaining problem described in Section 2.3, and satisfy* (3.4)-(3.6)*;*
3. *Stationarity: given (iii), the joint distribution $\Phi(\mathbf{x}, \xi)$ satisfies the inflow-outflow balance condition* (3.15)*, and investors' measures sum to one;*
4. *Market clearing: all asset markets clear, such that* (3.16) *is satisfied.*

In the following proposition, I show that the stationary equilibrium defined above exists and is unique. The proof can be found in Appendix A.3.

Proposition 3. *There exists a unique stationary equilibrium in the economy.*

I now proceed to solutions. Let $\mu(x^i_{\xi'}, \xi)$ denote the steady-state measure of investors with asset i holding $x^i_{\xi'}$ and patience type ξ, where $\xi', \xi \in \{h, l\}$. For instance, $\mu(x^1_h, h)$ is the steady-state measure of investors with patience type h, and holding x_h^1 units of asset 1 (while there are 2^{N-1} different combinations of other asset holdings). In the steady state, the stationarity condition (3.15) can be rewritten as

$$0 = -\zeta_{hl}\mu(x_h^i, h) + \zeta_{lh}\mu(x_h^i, l) + \lambda^i \mu(x_l^i, h) \tag{3.17}$$

$$0 = -\zeta_{hl}\mu(x_l^i, h) + \zeta_{lh}\mu(x_l^i, l) - \lambda^i \mu(x_l^i, h) \tag{3.18}$$

$$0 = -\zeta_{lh}\mu(x_h^i, l) + \zeta_{hl}\mu(x_h^i, h) - \lambda^i \mu(x_h^i, l) \tag{3.19}$$

$$0 = -\zeta_{lh}\mu(x_l^i, l) + \zeta_{hl}\mu(x_l^i, h) + \lambda^i \mu(x_h^i, l) \tag{3.20}$$

for all $i \in \{1, 2, \ldots, N\}$. Take equation (3.17) for example. This is the inflow-outflow balance condition for investors with patience type h and asset i holding x_h^i. The first term on the right-hand side represents outflow from the investor population, due to switching of patience type from type h to type l. The second term on the right-hand side represents inflow due to switching of patience type. The last term on the right-hand side represents inflow into the investor population due to trading by investors with patience type h but sub-optimal asset i holding. Although there is trading in other assets, it simply leads to flows among type h investors holding x_h^i units of asset i, and does not result in inflow into or outflow from this population. Because investors' measures sum to one, then

$$\mu(x_h^i, h) + \mu(x_h^i, l) + \mu(x_l^i, h) + \mu(x_l^i, l) = 1 \tag{3.21}$$

The steady-state measures, $\mu(x_h^i, h)$, $\mu(x_h^i, l)$, $\mu(x_l^i, h)$ and $\mu(x_l^i, l)$ are determined by the inflow-outflow balance conditions (3.17)-(3.20), and that investors' measures sum to one as given by (3.21). As shown in Appendix A.4, the measures are given by

$$\mu(x_h^i, l) = \mu(x_l^i, h) = \frac{\frac{\zeta_{hl}\zeta_{lh}}{\zeta_{hl}+\zeta_{lh}}}{\zeta_{lh} + \zeta_{hl} + \lambda^i} \tag{3.22}$$

$$\mu(x_h^i, h) = 1 - \pi - \mu(x_l^i, h) \quad \text{and} \quad \mu(x_l^i, l) = \pi - \mu(x_h^i, l) \tag{3.23}$$

for all $i \in \{1, 2, \dots, N\}$. Let X^i denote the total asset i holdings by investors. Then

$$X^i = [\mu(x_h^i, h) + \mu(x_h^i, l)]x_h^i + [\mu(x_l^i, h) + \mu(x_l^i, l)]x_l^i = \frac{\bar{D} - (r + \pi\epsilon + \Delta^i)\bar{P}^i}{\gamma\sigma^2} \tag{3.24}$$

where

$$\Delta^i = \frac{\zeta_{hl}\zeta_{lh}\epsilon}{\zeta_{hl} + \zeta_{lh}} \frac{\lambda^i(\theta_h - \theta_l) - \epsilon}{(r + \epsilon + \zeta_{lh} + \zeta_{hl} + \lambda^i\theta_l)(r + \zeta_{lh} + \zeta_{hl} + \lambda^i\theta_h)} \tag{3.25}$$

The wedge Δ^i modifies investors' aggregate demand for the asset. Compared to the Walrasian benchmark, investors' aggregate asset i demand in the economy is lower by $\frac{\Delta^i \bar{P}^i}{\gamma\sigma^2}$. When $\Delta^i = 0$, the aggregate asset demand coincides with that under the Walrasian benchmark. It is the case when, for example, $\lambda^i \to \infty$. If the demand wedge is positive $\Delta^i > 0$, X^i is lower than that in the benchmark case. If the demand wedge is negative $\Delta^i < 0$, X^i is higher than that under the benchmark.

3.5 Discussion

Before assessing the equilibrium properties, I wish to briefly discuss the importance of type-dependent bargaining powers of investors in this model. The existing literature on OTC markets generally assumes that investors' bargaining powers are constant.[9] Those models typically take an axiomatic approach to bargaining, and assume that trading parties Nash bargain and divide the joint trade surplus according to exogenously specified bargaining powers. In this model, investors' bargaining powers depend on their patience types. Importantly, the type-dependency of investor bargaining powers is an equilibrium outcome from strategic bargaining between investors and dealers, rather than an assumption. It thus marks a key departure from the literature.

To facilitate comparison with existing models, I shut down strategic bargaining in this model such that investors have a constant bargaining power $\theta \in (0, 1)$ when trading with dealers. That is, I set $\theta_h = \theta_l \equiv \theta$. Notice that when ϵ is relatively small, the first-order approximation of (3.25)

[9] For example, Duffie et al. (2005, 2007), Lagos and Rocheteau (2009), Afonso and Lagos (2015), Atkeson et al. (2015), Üslü (2019), Hugonnier et al. (2020) among others.

is given by

$$\Delta^i \doteq \frac{\zeta_{hl}\zeta_{lh}\epsilon}{\zeta_{hl}+\zeta_{lh}} \frac{\lambda^i(\theta_h-\theta_l)}{(r+\epsilon+\zeta_{lh}+\zeta_{hl}+\lambda^i\theta_l)(r+\zeta_{lh}+\zeta_{hl}+\lambda^i\theta_h)} \tag{3.26}$$

Under the condition that $\theta_h = \theta_l$, the demand wedge $\Delta^i \doteq 0$ and investors' aggregate asset i holdings $X^i \doteq \frac{\bar{D}-(r+\pi\epsilon)\bar{P}^i}{\gamma\sigma^2}$.[10] In this case, asset i's inter-dealer clearing price $\bar{P}^i$ first-order converges to the price under the Walrasian benchmark. That is, when investors are not subject to holding restrictions, asset illiquidity can have large implications for portfolio holdings, but little impact on price level. Intuitively, when an asset's liquidity increases, type h investors' demand for the asset increases while type l investors' demand for the asset decreases. Hence, the aggregate demand is little affected, leading to hardly any impact on price.[11] However, the above result no longer holds if $\theta_h \neq \theta_l$. As discussed in the next section, when strategic bargaining leads to type-dependency of investor bargaining powers, asset liquidity affects both portfolio choice and asset prices.

4 Equilibrium Properties

In this section, I provide the equilibrium properties. I show how asset liquidity affects portfolio choice and asset pricing, and discuss the intuition behind these results. I also document the model implications for transaction costs, which have been the main focus of the literature on OTC markets and microstructure. The results in this section give rise to several novel predictions, that are testable with data.

4.1 Portfolio Choice

Proposition 2 suggests that an investor's optimal demand for an asset depends on the asset's liquidity. As explained before, when an asset is illiquid, investors attenuate their demand (relative

[10] In this model, investors experience idiosyncratic shocks to their discount rates, which enter the marginal value functions non-linearly. The second-order term in (3.25) results from the non-linear impact of discount rates.

[11] This result is consistent with Garleanu (2009), who shows that asset price in a search market is the same as that in a Walrasian market, and the price is independent of the asset's liquidity. As in this paper, Garleanu (2009) relaxes the $\{0, 1\}$ binary holding restriction typical in the OTC literature. He shows that the positive relation between price and liquidity documented in prior studies (e.g., Duffie et al. (2005, 2007), Weill (2008)) largely result from binding portfolio constraints.

to the Walrasian benchmark) towards the market average. Their demand attenuations are greater when the asset is more illiquid. Since $\lambda^1 > \lambda^2 > \cdots > \lambda^N$, Proposition 2 implies that

$$x_h^1 > x_h^2 > \cdots > x_h^N \quad \text{and} \quad x_l^1 < x_l^2 < \cdots < x_l^N \tag{4.1}$$

That is, patient investors tilt their portfolios towards the more liquid assets, while impatient investors hold more of the illiquid assets. When a patient investor experiences a patience shock and becomes impatient, they wish to sell more of the liquid assets and less of the illiquid assets. This helps explain why during market stress when investors' aggregate liquidity needs are high, investors tend to sell liquid assets more aggressively. The observation that selling pressure concentrates in more liquid assets during crises (e.g., Ma et al. (2022)) results from investors' dynamic portfolio choice. Moreover, from (3.25) and (3.26), it is immediate that for small enough ϵ, the demand wedge $\Delta^i > 0$ for all i. In other words, asset illiquidity results in lower aggregate demand for the asset, relative to the Walrasian benchmark. The positive demand wedge arises because patient investors attenuate their demand more than impatient investors, due to strategic bargaining.

In equilibrium, investors with sub-optimal holdings wish to trade. Type l investors holding x_h^i units of asset i wish to sell their holdings down to x_l^i, while type h investors holding x_l^i units of asset i wish to increase their holdings to x_h^i. The intensity of trading is λ^i. Hence, the instantaneous trading volume in asset i is $\mathcal{V}^i = \lambda^i \big[\mu(x_h^i, l) + \mu(x_l^i, h)\big] \big|x_h^i - x_l^i\big|$. Using (3.22), I provide the expression for $\mathcal{V}^i$ below.

Proposition 4. *The instantaneous trading volume by investors in asset i is*

$$\mathcal{V}^i = \frac{2\zeta_{hl}\zeta_{lh}}{\zeta_{hl} + \zeta_{lh}} \frac{\lambda^i}{\zeta_{lh} + \zeta_{hl} + \lambda^i} \big|x_h^i - x_l^i\big| \qquad \forall\, i \in \{1, 2, \ldots, N\} \tag{4.2}$$

The trading volume $\mathcal{V}^i$ increases in the asset's liquidity λ^i. Thus, more liquid assets tend to have higher trading volumes. On the extensive margin, investors are more likely to find dealers to trade when an asset is more liquid. On the intensive margin, because investors hold more extreme positions in more liquid assets, their desired trade quantities are larger when an asset is more liquid.

4.2 Asset Prices

A key goal of the paper is to study how asset prices depend on asset liquidity in the cross-section. In the model, I assume that the assets have i.i.d. dividend flows and the same fixed supply, and differ only in their liquidity as captured by their respective meeting intensities. Investors' aggregate asset holdings and market clearing jointly pin down the central inter-dealer clearing price of an asset.

Proposition 5. *The market clearing price of asset i in the inter-dealer market is*

$$\bar{P}^i = \frac{\bar{D} - \gamma\sigma^2 s}{r + \pi\epsilon + \Delta^i} \tag{4.3}$$

for all $i \in \{1, 2, \ldots, N\}$, and Δ^i is given by (3.25).

The derivations for Proposition 5 can be found in Appendix A.4. Throughout the paper, I measure the price level of an asset using the asset's inter-dealer clearing price. The idea is that in semi-centralized markets intermediated by dealers, dealers effectively trade in the central inter-dealer market on the investors' behalf at the prevailing inter-dealer price. Hence, the inter-dealer price can be thought of as the price at which the market clears. The difference between the inter-dealer clearing price and an investor-dealer trading price thus represents a transaction cost incurred by the investor in exchange for the dealer's intermediation services.

Consider two assets i and j with $\lambda^i > \lambda^j$ (i.e., asset i is more liquid than asset j), which asset should trade at higher price? From (4.3), whether $\bar{P}^i$ is higher or lower than $\bar{P}^j$ depends on the relative sizes of the two assets' demand wedges Δ^i and Δ^j. If $\Delta^i > \Delta^j$, then $\bar{P}^i < \bar{P}^j$; if $\Delta^i < \Delta^j$, then $\bar{P}^i > \bar{P}^j$. From (3.25), an asset's demand wedge is a function of the asset's liquidity (or meeting intensity for that asset). Let λ denote an asset's meeting intensity with $\lambda \in \{\lambda^1, \lambda^2, \ldots, \lambda^N\}$. The asset's demand wedge can be written as a function of λ, that is,

$$\Delta(\lambda) = \frac{\zeta_{hl}\zeta_{lh}\epsilon}{\zeta_{hl} + \zeta_{lh}} \frac{\lambda(\theta_h - \theta_l) - \epsilon}{(r + \epsilon + \zeta_{lh} + \zeta_{hl} + \lambda\theta_l)(r + \zeta_{lh} + \zeta_{hl} + \lambda\theta_h)} \tag{4.4}$$

Thus, whether $\Delta(\lambda)$ is increasing or decreasing in λ is of interest. If $\Delta'(\lambda) < 0$, the demand wedge is decreasing in asset liquidity, and the price is increasing in asset liquidity. By contrast, if $\Delta'(\lambda) > 0$, the demand wedge is increasing in asset liquidity, and the price is decreasing in asset liquidity.

Lemma 6. *$\Delta'(\lambda) > 0$ if $\lambda < \bar{\lambda}$, and $\Delta'(\lambda) < 0$ if $\lambda > \bar{\lambda}$, where*

$$\bar{\lambda} = \frac{\epsilon\theta_l\theta_h + \sqrt{\epsilon^2\theta_l^2\theta_h^2 + (\theta_h - \theta_l)\theta_l\theta_h[(\theta_h - \theta_l)k_1k_2 + \epsilon(k_1\theta_h + k_2\theta_l)]}}{(\theta_h - \theta_l)\theta_l\theta_h} \tag{4.5}$$

with $k_1 = r + \epsilon + \zeta_{lh} + \zeta_{hl}$ and $k_2 = r + \zeta_{lh} + \zeta_{hl}$. Moreover, $\Delta''(\lambda) < 0$.

The proof for Lemma 6 is shown in Appendix A.5. The lemma states that the effect of asset liquidity on the asset's demand wedge, and thus asset price, is non-monotonic. When an asset is sufficiently liquid ($\lambda > \bar{\lambda}$), increasing the asset's liquidity raises its price. However, when the asset is highly illiquid ($\lambda < \bar{\lambda}$), increasing the asset's liquidity reduces its price instead.

4.2.1 Intuition

I now explain the intuition underlying Lemma 6. As explained before, although both patient and impatient investors attenuate their demand, patient investors attenuate more than impatient investors.[12] Demand attenuations result from investors' desire to hedge against possible holding imbalances due to asset illiquidity and inability to trade. Asymmetry in demand attenuations by patient and impatient investors stems from strategic bargaining, as patient investors further lower their asset demand to internalize the effect of weaker bargaining powers when they are in the shock state. The demand wedge is due to the asymmetric demand attenuations by investors, which are caused by strategic bargaining.

I start by considering the limit case with $\lambda \to 0$. In this case, the asset is perfectly illiquid, and it would take an investor infinite amount of time to meet a dealer. Trading, and thus bargaining, never takes place in this limit. Because demand wedge results from strategic bargaining, and

[12] Demand attenuations are relative to the Walraisian benchmark demand. Patient investors decrease their demand relative to the benchmark, while impatient investors increase their demand relative to the benchmark.

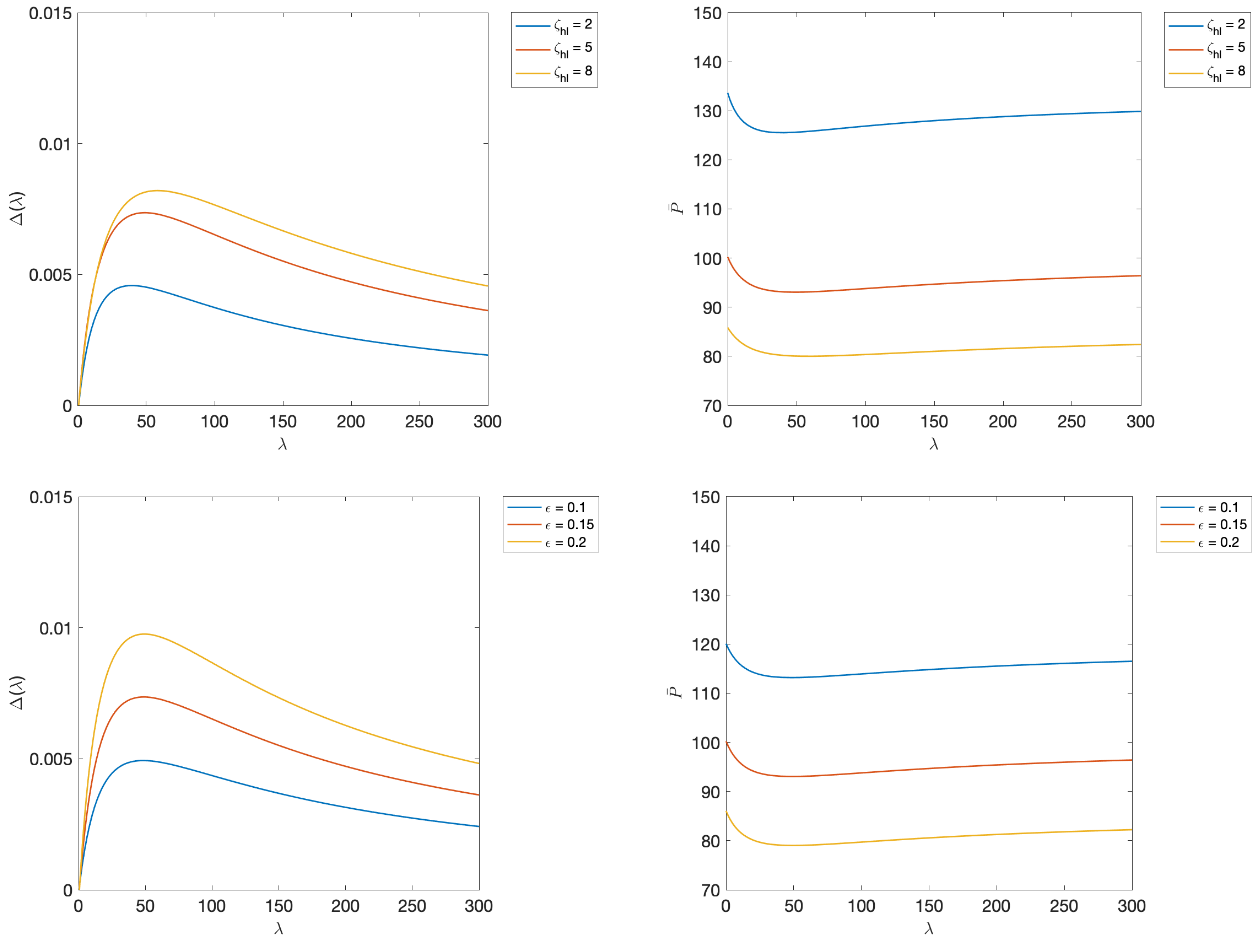
bargaining is irrelevant in this limit case, there is no demand wedge with $\lambda \to 0$. For a small increase in λ, the effect of bargaining becomes pertinent, and a positive demand wedge emerges (since patient investors attenuate their demand more than impatient investors), leading to a decrease in the asset price. However, an increase in λ also decreases demand attenuations by both patient and impatient investors. This is because investors attenuate their asset demand and hold less extreme positions to hedge against future holding imbalances. This hedging motive decreases when the asset becomes more liquid and easier to trade. Thus, as λ continues to increase, the demand wedge shrinks and the asset price rises. As $\lambda \to \infty$ (that is, the asset becomes perfectly liquid), investors no longer feel the need to attenuate their asset demand, and the demand wedge again converges to zero.

In other words, an increase in the asset's liquidity λ has two countervailing effects. First, an increase in asset liquidity makes the aforementioned bargaining consideration more salient, because investors are more likely to come across dealers and engage in bargaining. The amplification of the bargaining consideration tends to result in greater asymmetry in demand attenuations by patient and impatient investors, leading to larger demand wedge and lower asset price. This is the *bargaining* effect, and is captured by the numerator of the second fraction term on the right-hand side of (4.4). Second, an increase in asset liquidity makes it easier to trade the asset, leading to lower demand attenuations by both patient and impatient investors. The lower demand attenuations tend to decrease the demand wedge (which is a weighted average of the signed demand attenuations by investors) and increase the asset price. This is the *demand attenuations* effect, and is captured by the denominator of the second fraction term on the right-hand side of (4.4). When λ is low, the former effect dominates and asset price decreases in asset liquidity. When λ is sufficiently high, the latter effect dominates and asset price increases in asset liquidity.

4.2.2 Numerical Example

I illustrate the results from Lemma 6 with a simple numerical example, shown in Figure 2. The baseline parameter values are $\bar{D} = 10$, $\sigma = 0.1$, $s = 0.2$, $r = 0.05$, $\gamma = 5$, $z = 0.5$, $\zeta_{lh} = 10$, $\zeta_{hl} = 5$ and $\epsilon = 0.15$.

Figure 2: Numerical Example



Note: The figure plots the demand wedge $\Delta(\lambda)$ and the inter-dealer clearing price $\bar{P}$ against the meeting intensity λ, for various values of ζ_{hl} and ϵ. The baseline parameter values are as follows: $\bar{D} = 10$, $\sigma = 0.1$, $s = 0.2$, $r = 0.05$, $\gamma = 5$, $z = 0.5$, $\zeta_{lh} = 10$, $\zeta_{hl} = 5$ and $\epsilon = 0.15$. The top two panels plots the functions for $\zeta_{hl} = 2, 5, 8$. The bottom two panels plots the functions for $\epsilon = 0.1, 0.15, 0.2$.

The switching intensities $\zeta_{hl} = 5$ and $\zeta_{lh} = 10$ imply that a patient investor remains patient for 0.2 years on average, while an impatient investor recovers from their shock within 0.1 years on average. Dealers and patient investors discount time at 5% per annum, which roughly matches the risk-free rate in the U.S. Impatient investors discount time at 20%, reflecting their imminent liquidity needs. The supply of asset is 0.2, a fraction of the total measure of investors. Investors' risk aversion coefficient is 5, which is commonly used in the asset pricing literature. Dealers' offer probability, which captures their market powers or bargaining skills outside the model, is 0.5.

Under the baseline parameter values, the bargaining powers of patient and impatient investors are $\theta_h = 0.5$ and $\theta_l = 0.2$ respectively. In Figure 2, the orange lines (in the middle) plot the demand wedge $\Delta(\lambda)$ against the asset liquidity λ under the baseline parameter values. The liquidity threshold $\bar{\lambda}$ at the inflection point is roughly 49 in the base case. A meeting intensity of 49 means that it takes approximately 5 trading days to successfully transact in the given asset.[13]

The top two panels of Figure 2 plot the demand wedge and the clearing price against the asset liquidity, for various values of ζ_{hl}. The bottom two panels plot the demand wedge and the clearing price against the asset liquidity, for various values of ϵ. The liquidity threshold $\bar{\lambda}$ appears to increase in both ζ_{hl} and ϵ.

4.2.3 Asset Liquidity and Prices

Proposition 7. *Suppose $\lambda^1 > \cdots > \lambda^i > \bar{\lambda} > \lambda^{i+1} > \cdots > \lambda^N$. Then, among the first i assets, $\bar{P}^1 > \cdots > \bar{P}^i$; among the last $N - i$ assets, $\bar{P}^{i+1} < \cdots < \bar{P}^N$.*

Proposition 7 follows directly from Lemma 6, and provides the key result of the model. That is, the relationship between asset prices and asset liquidity is non-monotonic in the cross-section. Specifically, the price-liquidity relationship is positive for assets with meeting intensities above the liquidity threshold $\bar{\lambda}$, but negative for assets with meeting intensities below $\bar{\lambda}$. In other words, among sufficiently liquid assets, relatively more liquid assets trade at higher prices. However, among highly illiquid assets, relatively more liquid assets trade at lower prices instead.

The proposition above makes a cross-sectional statement. However, notice that the liquidity threshold $\bar{\lambda}$ depends on other model parameters. As the numerical example illustrates, under appropriate parameter values, $\bar{\lambda}$ increases in both ζ_{hl} and ϵ. Since ζ_{hl} measures how likely investors are to experience patience shocks and ϵ captures the magnitude of such shocks, large values of ζ_{hl} and ϵ are often associated with times of stress when investors' liquidity needs are high. Hence, the negative relationship between asset prices and liquidity is more prevalent during market stress. The average price-liquidity relationship may become negative during severe crises.

[13] In practice, many OTC assets take considerably longer time to trade. For example, many corporate bonds often take days if not weeks to transact.

4.3 Transaction Costs

The inter-dealer price $\bar{P}^i$ of asset i is the market clearing price of the asset, and has been my focus so far. However, due to the bilateral nature of investor-dealer trading, investors' actual transaction prices differ from the clearing price. The difference between an investor's trade price and the inter-dealer price can be viewed as the transaction cost paid by the investor in exchange for the dealer's intermediation services. Let $Cost^{i,sell}$ denote the transaction cost incurred by a selling investor of asset i, and let $Cost^{i,buy}$ denote the transaction cost incurred by a buying investor of asset i. The transaction costs $Cost^{i,sell}$ and $Cost^{i,buy}$ are defined as the percentage deviation of the trade prices from the inter-dealer clearing price. That is,

$$Cost^{i,sell} = \frac{\bar{P}^i - P_l^i(\mathbf{x})|_{x^i=x_h^i}}{\bar{P}^i} \quad \text{and} \quad Cost^{i,buy} = \frac{P_h^i(\mathbf{x})|_{x^i=x_l^i} - \bar{P}^i}{\bar{P}^i} \tag{4.6}$$

since in equilibrium, type l investors holding x_h^i units of asset i wish to sell, and type h investors holding x_l^i units of asset i wish to buy. As shown in Appendix A.6, the transaction costs are given by the following proposition.

Proposition 8. *The transaction cost paid by selling investors of asset i and the transaction cost paid by buying investors of asset i, where $i \in \{1, 2, \ldots, N\}$, are*

$$Cost^{i,sell} = \frac{(1-\theta_l)\epsilon}{2(r+\epsilon+\zeta_{lh}+\zeta_{hl}+\lambda^i\theta_l)} \tag{4.7}$$

$$Cost^{i,buy} = \frac{(1-\theta_h)\epsilon}{2(r+\zeta_{lh}+\zeta_{hl}+\lambda^i\theta_h)} \tag{4.8}$$

Proposition 8 suggests that transaction costs decrease in asset liquidity. That is, more liquid assets have lower transaction costs. Hence, although the relationship between asset prices and asset liquidity is non-monotonic, transaction costs are monotonically decreasing in liquidity. The proposition thus provides justification for commonly used liquidity measures such as transaction costs and bid-ask spreads. The model also has the potential to rationalize the empirical pattern that transaction cost and trade size are negatively related (e.g., Edwards et al. (2007)), since transaction cost decreases in asset liquidity while trade size increases in asset liquidity. Moreover, and new to the literature, the proposition implies that transaction costs are asymmetric depending on the trade direction. Given a relatively small ϵ, selling investors incur higher transaction costs than buying investors, when trading with dealers.

4.4 Empirical Predictions

I summarize the model's main empirical predictions that are novel to the literature. I then test these predictions empirically in Section 5.

Prediction 1a. *Among sufficiently liquid assets, more liquid assets tend to trade at higher prices. Among highly illiquid assets, more liquid assets tend to trade at lower prices instead. The effect of asset liquidity on prices is generally increasing in the liquidity level.*

The above prediction makes a cross-sectional statement about the non-monotonic relationship between asset prices and liquidity. Specifically, the price-liquidity relationship is positive among sufficiently liquid assets, but negative among highly illiquid assets. Furthermore, an increase in asset liquidity exerts more positive impact on price for assets that are more liquid already. This result follows from the convexity of prices with respect to asset liquidity.

Prediction 1b. *The negative relationship between asset prices and asset liquidity is more prevalent during market stress. In severe crises, the average price-liquidity relationship may become negative.*

During market stress when investors have greater liquidity needs, the liquidity threshold (below which the price-liquidity relationship is negative) is higher. Hence, more assets tend to exhibit negative price-liquidity relationship. When market stress is severe, the liquidity threshold may be high enough such that on average the price-liquidity relationship is negative in the market.

Prediction 2. *Transaction costs are higher for selling investors than for buying investors.*

Compared to buying investors, selling investors are more impatient and have weaker bargaining powers, resulting in higher transaction costs incurred by them when trading with dealers.

5 Empirical Evidence

In this section, I use the U.S. corporate bond market as the empirical setting and test the model predictions. Corporate bonds in the U.S. are traded over-the-counter and intermediated by dealers. The market has a two-tiered structure: Investors must trade with dealers bilaterally through sequential search and bargaining, while dealers can also trade with each other in a more centralized inter-dealer market. The corporate bond market has garnered significant research interests, given its large size and significance in firms' financing decisions.[14] Transaction-level data are available for U.S. corporate bonds, through FINRA's Trade Reporting and Compliance Engine (TRACE).

5.1 Data

The main dataset used for my empirical analysis is the enhanced version of TRACE, which is provided by the Financial Industry Regulatory Authority (FINRA). The enhanced TRACE dataset contains detailed trade-level information on U.S. corporate bond transactions, including bond CUSIP, trade price and quantity, trade execution date and time, a counterparty identifier that separates inter-dealer trades from dealer-customer trades, a dealer buy/sell indicator which specifies whether a trade is a dealer buy or sell, and a trading market indicator that shows whether a trade is a primary market transaction or a secondary market transaction.

For bonds included in the TRACE data, I use the Mergent Fixed Income Securities Database (FISD) to obtain information on the characteristics of each bond, including issuance and maturity dates, issue amount, and credit ratings. For corporate bonds' credit ratings, the FISD dataset contains a complete history of rating changes by each of the three major rating agencies: Standard & Poor's (S&P), Moody's, and Fitch. To construct a credit rating measure for each bond on each day, I use the median rating as the composite credit rating. Specifically, I assign a numeric value to each notch of S&P ratings, with 1, 2, 3, 4, ... denoting AAA, AA+, AA, AA-, ..., respectively. I then follow the same approach and assign numeric values for Moody's and Fitch ratings. A bond's

[14]The U.S. corporate bond market is estimated to be over $10 trillion in size as of Q3 2022, representing over 20% of the total fixed income market in the U.S. (Source: SIFMA).

composite rating on a given day is the median of its ratings on that day. Table 1 documents the credit rating scales and associated numeric values by the three rating agencies. Bonds with composite ratings below 11 are considered investment-grade (IG), otherwise they are high-yield (HY).

Table 1: Rating Scales

Moody's	S&P	Fitch	Value
Aaa	AAA	AAA	1
Aa1	AA+	AA+	2
Aa2	AA	AA	3
Aa3	AA-	AA-	4
A1	A+	A+	5
A2	A	A	6
A3	A-	A-	7
Baa1	BBB+	BBB+	8
Baa2	BBB	BBB	9
Baa3	BBB-	BBB-	10
Ba1	BB+	BB+	11
Ba2	BB	BB	12
Ba3	BB-	BB-	13
B1	B+	B+	14
B2	B	B	15
B3	B-	B-	16
Caa1	CCC+	CCC+	17
Caa2	CCC	CCC	18
Caa3	CCC-	CCC-	19
Ca	CC	CC	20
C	C	C	21
	D	DDD	22
		DD	23
		D	24

Note: This table reports the credit rating scales and associated numeric values assigned by the three major credit rating agencies, namely Moody's, S&P and Fitch. Credit ratings data for corporate bonds are obtained from the Mergent Fixed Income Securities Database (FISD). A numeric value below 11 is considered investment-grade (IG) while a numeric value equal or above 11 is considered non-investment grade or high-yield (HY).

I obtain bond transaction data from TRACE for the period between January 1, 2005 and December 31, 2021. I follow the literature and apply standard filtering procedures (e.g., Dick-Nielsen (2014)) to clean the TRACE datasets. Following the literature, I also restrict my sample to

U.S. corporate debentures that have fixed coupons and are not convertible, putable, asset-backed, exchangeable, privately-placed, perpetual, or preferred securities. I further exclude any secured lease obligations and bonds quoted in foreign currencies. I then delete observations whose trade dates occur on or after the bond maturity dates, and observations with non-positive reported prices. To focus on secondary market trades, I exclude trades with trading market indicator "P1" which are primary market transactions. The ending sample contains more than 119 million transactions in 23,371 unique corporate bonds issued by 4,920 different issuers.

Price Measure. The TRACE dataset reports transaction prices for all investor-dealer and inter-dealer corporate bond transactions. Guided by the model, I calculate the price for each bond on each day by averaging the inter-dealer prices of the bond on that day. To facilitate comparison across bonds, I compute the yield-to-maturity for each bond on each day based on the price, and then calculate the bond-day level credit spread by subtracting the yield of the corresponding Treasury security.[15]

Liquidity Measure. My main measure for bond liquidity is transaction cost, which is a commonly used liquidity measure in the literature and justified by the model results. I estimate the transaction cost for each investor-dealer trade k as

$$Cost_k = \ln\left(\frac{Trade\ Price_k}{Benchmark\ Price_k}\right) \times Trade\ Direction_k \tag{5.1}$$

where $Trade\ Price_k$ is the reported price for bond trade k, and $Benchmark\ Price_k$ is the most recent inter-dealer trade price prior to trade k. $TradeDirection_k$ takes the value $+1$ for an investor purchase and -1 for an investor sale. I then multiply $Cost_k$ by 10,000 to get the transaction cost in basis points.[16] To limit noise, I follow the literature and exclude retail-sized trades (that is, trades with size below \$100,000) from the calculation. The bond-day level transaction cost measures are obtained by averaging transaction costs for each bond on each day.

[15]The Treasury yield data are from the Gürkaynak et al. (2007) database available through the Federal Reserve, with linear interpolations between provided maturities when necessary.

[16]The transaction cost measure follows the literature (e.g., Hendershott and Madhavan (2015) and O'Hara and Zhou (2021)) and is a direct analog of how transaction costs are defined in the model.

5.2 Asset Prices and Liquidity

I start by examining the relationship between corporate bond prices (measured by credit spreads) and liquidity (measured by transaction costs) in the cross-section. To eliminate differences across bonds due to heterogeneity in collateral values and seniority rankings, I restrict my attention to senior unsecured notes only. To mitigate noise due to outliers, I trim the top and the bottom 1% of the credit spread measures from my analysis. I also exclude observations with time-to-maturity below one year and observations with missing credit ratings, to be consistent with the literature.

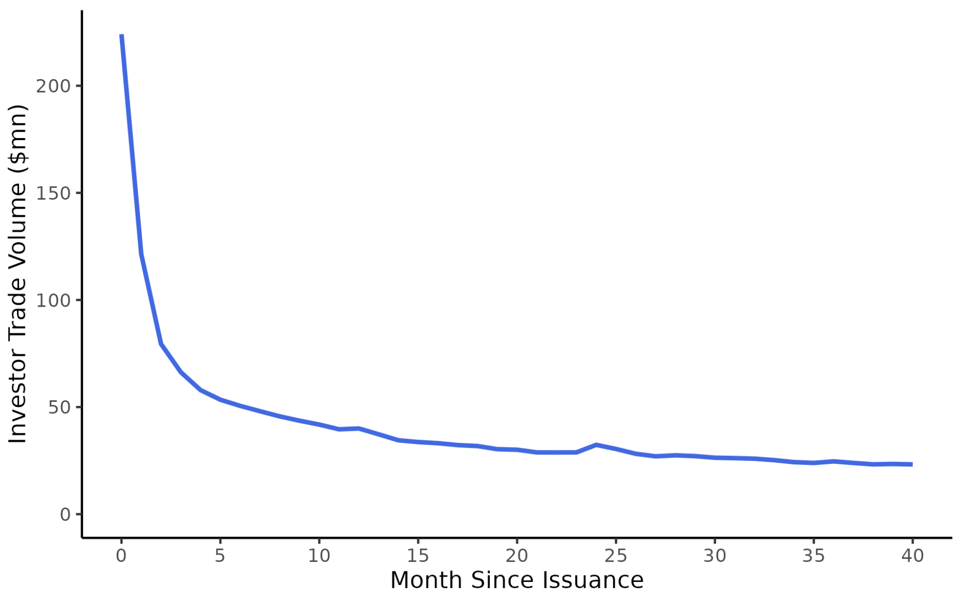
The theory predicts that the credit spreads of sufficiently liquid bonds are positively related to transaction costs, while the credit spreads of illiquid bonds are negatively related to transaction costs. However, regressing credit spreads directly on transaction costs may be problematic due to endogeneity concerns. One challenge is omitted variables. Transaction costs are potentially related to price measures through factors other than bond liquidity, such as bond characteristics, issuer credit risk, and market conditions. The omitted variables bias can be mitigated by including appropriate controls. Another challenge is simultaneity. Transaction costs are calculated using prices, thus creating a bidirectional and simultaneous relationship between them.

To address these issues, I adopt a two-stage least squares (2SLS) approach and rely on an instrument that exploits plausibly exogenous variation in bond liquidity. In particular, I exploit the institutional feature that newly-issued bonds tend to be more liquid than older bonds issued by the same issuer.[17] The idea is that when new bonds are issued, investors actively trade them until they are gradually absorbed into the holdings of long-term investors. Because long-term investors tend to have longer time horizons and trade less frequently, trading volume and liquidity decrease with bond age. Figure 3 plots the decay in trading volume as bond age increases. The decline in volume is particularly steep during the first 12 months or so from issuance. Hence, I construct an instrumental variable based on whether a bond is newly-issued within the past year. The instrument is a dummy variable $New\ Bond_{it}$ which equals one if bond i's time from issuance

[17] E.g., see Sarig and Warga (1989), Houweling et al. (2005), Edwards et al. (2007), Mizrach (2015).

is below one year as of day t, and zero otherwise.[18] Table 2 reports the summary statistics of newly-issued and older bonds respectively.

Figure 3: Monthly Trading Volume from Issuance



Note: The figure plots the average monthly investor-dealer trading volume in the issuance month and forward, using corporate bond transaction data from the enhanced TRACE and bond characteristics data from the Mergent FISD for the period between 1/1/2005 and 12/31/2021.

A valid instrument must satisfy both the relevance condition and the exclusion restriction. The discussion above supports that the instrument $New\ Bond_{it}$ satisfies the relevance condition. Figure 4 provides further evidence by showing that newly-issued bonds have lower transaction costs and higher trading volumes than older bonds. Furthermore, the exclusion restriction is plausibly satisfied, as bond age is unlikely to directly affect bond yield or credit spread beyond its impact on liquidity, particularly if potential confounding factors such as fundamental risks, bond characteristics and market conditions are controlled for.

To test whether the price-liquidity relationship is non-monotonic in the cross-section of corporate bonds, I conduct split-sample analysis. I divide the full estimation sample into sub-samples containing relatively liquid bonds and illiquid bonds respectively, based on prior month's transaction costs, credit ratings, and time-to-maturity. I then estimate the following first-stage

[18] Here, I define a newly-issued bond as one issued within the past year (i.e., the bond's age is below one year). The one-year bond age cutoff is based on the result presented in Figure 3. In untabulated results, using alternative bond age cutoffs (e.g., 90 days, 2 years or 3 years) yields similar results.

Table 2: Bond Characteristics of New Bonds and Old Bonds

	New Bonds Num of Obs ($n = 1,560,158$)				Old Bonds Num of Obs ($n = 8,149,719$)			
	Mean	Q1	Median	Q3	Mean	Q1	Median	Q3
Price	101.7	98.8	101.0	104.4	105.2	99.9	104.3	110.1
Credit Spread	185.1	77.6	125.6	224.1	234.0	81.3	144.7	281.1
Transaction Cost	17.5	0.0	9.4	27.7	23.5	0.0	10.3	34.7
Trading Volume	6.8	0.2	1.2	6.1	2.2	0.0	0.2	1.3
Rating	8.0	6.0	8.0	10.0	9.0	7.0	9.0	10.0
Time-to-Maturity	10.2	4.8	9.1	9.8	8.6	3.3	5.7	8.8
Bond Age	0.5	0.2	0.5	0.7	5.7	2.5	4.2	7.0
Issue Amount	984.3	500.0	750.0	1250.0	834.9	400.0	600.0	1000.0

Note: This table reports summary statistics for newly-issued bonds (i.e., bonds aged below one year) and older bonds, using the enhanced TRACE data and the Mergent FISD data from 1/1/2005 to 12/31/2021. Observations are at the bond-day level. *Price* represents the average inter-dealer price. *Credit Spread* is the corresponding credit spread expressed in basis points. *Transaction Cost* represents the average transaction cost estimated using Eq. (5.1), reported in basis points. *Trading Volume* is the gross daily investor-dealer transaction volume expressed in $millions. *Rating* is the numerical composite bond rating where $AAA = 1, AA+ = 2, AA = 3, \ldots$. *Time-to-Maturity* and *Bond Age* are time left until maturity and time since issuance, respectively, in years. *Issue Amount* is offering amount in $millions.

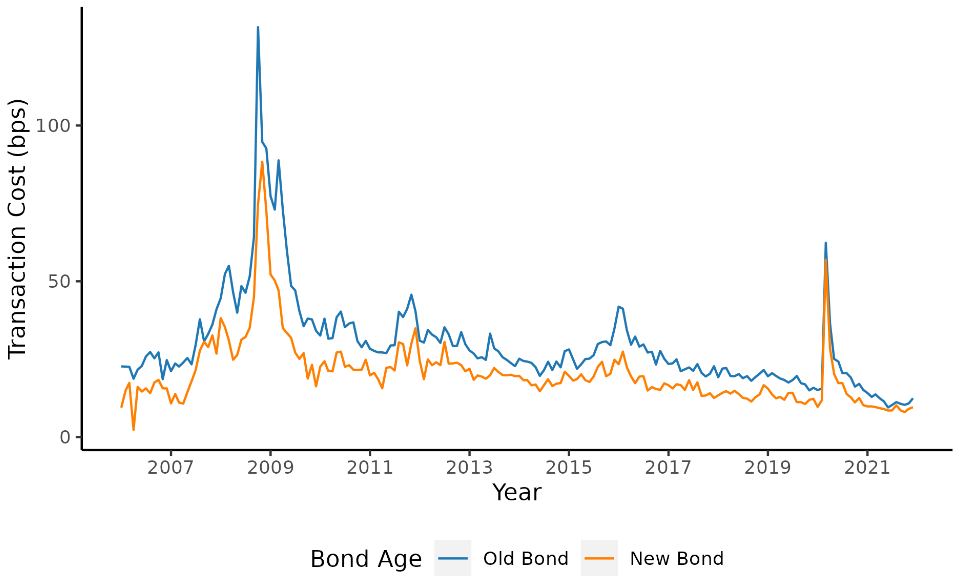
specification for each sub-sample

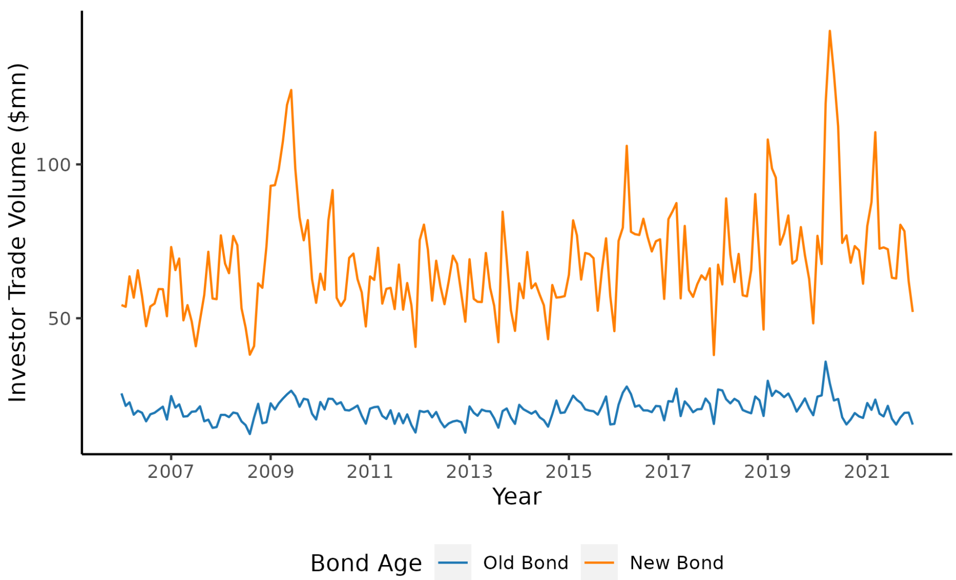
$$Transaction\ Cost_{it} = \phi New\ Bond_{it} + Controls + \alpha_{jt} + \epsilon_{it} \tag{5.2}$$

where i indexes bond, j indexes issuer, and t indexes day. $Transaction\ Cost_{it}$, denoting the transaction cost of bond i on day t, is the main liquidity measure. $Controls$ include bond-day level characteristics such as time-to-maturity and the logarithm of issue amount. The key controls are the issuer-day fixed effects α_{jt}, which account for time-varying fundamental differences at the issuer-level. The specification thus compares the same issuers' bonds on the same day, while controlling for differences due to maturities and issue sizes. Standard errors are clustered at the bond and the day levels.

The first-stage results are reported in Panel A of Table 3. The estimation sample is partitioned into sub-samples using cutoffs based on prior month's transaction costs, credit ratings, and time-to-maturity. Specifically, bonds whose transaction costs in the prior month are below the 75th percentile (Low Cost), bonds rated investment-grade (IG), and short-term bonds with

Figure 4: Transaction Cost and Trading Volume by Bond Age





Note: The figure plots the average monthly transaction cost and the average monthly investor-dealer trading volume from 1/1/2006 to 12/31/2021 for new bonds and old bonds respectively. New bonds are those with bond age below one year, and old bonds are those with bond age over a year. The sample is constructed using data from the enhanced TRACE and the Mergent FISD.

time-to-maturity below 3 years (ST) are considered relatively liquid. By contrast, bonds whose transaction costs in the prior month are above the 75th percentile (High Cost), bonds rated below investment-grade or high-yield (HY), and long-term bonds with time-to-maturity above 3 years (LT) are considered relatively illiquid.[19] Across all sub-samples, the transaction costs of newly-issued bonds (that is, $New\ Bond_{it} = 1$) are 3-6 bps, or 13-25%, lower than those of older bonds. All of the estimates are highly statistically significant.

By estimating the first-stage specification (5.2), I obtain the predicted values for the transaction cost measures $\widehat{Transaction}\ Cost_{it}$ for each sub-sample. In the second stage of the 2SLS analysis, I regress the outcome variable of interest, credit spreads, on these predicted values from the first stage. I estimate the following second-stage specification for each sub-sample

$$Credit\ Spread_{it} = \psi \widehat{Transaction}\ Cost_{it} + Controls + \alpha_{jt} + \epsilon_{it} \tag{5.3}$$

The coefficient of interest is ψ. Positive ψ means that more liquid bonds (i.e., bonds with lower transaction costs) trade at lower credit spreads or higher prices, while negative ψ means that more liquid bonds trade at higher credit spreads or lower prices instead. Panel B of Table 3 reports the second-stage regression results. Columns (1), (3) and (5) correspond to sub-samples of relatively liquid corporate bonds, while Columns (2), (4) and (6) correspond to sub-samples of relatively illiquid bonds, based on prior month's transaction costs, credit ratings and time-to-maturity. In the sub-samples containing relatively liquid bonds, the slope estimates on the instrumented transaction cost measures are always positive and highly statistically significant. By contrast, in the sub-samples of high-cost bonds and HY bonds, the slope estimates are negative and statistically significant at the conventional levels. In the sub-sample of long-term bonds, the slope estimate is still positive, but is much lower in magnitude compared to that of Column (5). In summary, the regression results presented so far confirm Prediction 1a and show that the relationship between prices and liquidity is non-monotonic in the cross-section.

In Figure 5, I plot the quarterly time series of the average credit spread differential between newly-issued bonds (that is, bonds aged one year or below) and matched old bonds. The matching procedure is such that each newly-issued bond is matched to the same issuer's old bonds (aged three years or above) with maturity dates no more than one year before or one year after the

[19] It is well documented that corporate bonds with higher credit ratings and shorter maturities tend to be more liquid (e.g., Edwards et al. (2007)).

Table 3: Effect of Transaction Costs on Credit Spreads

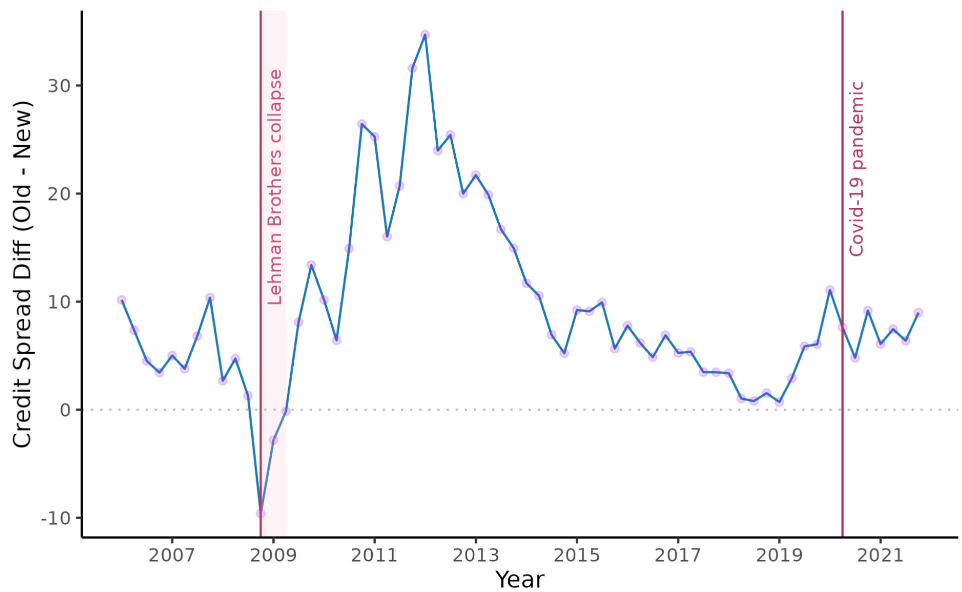
	Panel A: First Stage					
	Transaction Cost (bps)					
	(1)	(2)	(3)	(4)	(5)	(6)
New Bond	-3.061***	-6.026***	-4.631***	-2.852***	-2.529***	-5.415***
	(0.135)	(0.768)	(0.163)	(0.574)	(0.205)	(0.188)
Time-to-Maturity	0.520***	0.822***	0.681***	0.672***	2.383***	0.576***
	(0.017)	(0.123)	(0.016)	(0.205)	(0.138)	(0.031)
Issue Amount	-3.967***	-15.48***	-6.282***	-10.89***	-2.992***	-8.480***
	(0.165)	(1.036)	(0.226)	(1.133)	(0.244)	(0.366)
	Panel B: Second Stage					
	Credit Spread (bps)					
	(1)	(2)	(3)	(4)	(5)	(6)
Transaction Cost	0.424***	-0.502*	0.801***	-2.760***	2.538***	0.938***
	(0.156)	(0.297)	(0.094)	(1.017)	(0.368)	(0.097)
Time-to-Maturity	2.520***	2.570***	2.329***	3.866***	10.18***	1.658***
	(0.101)	(0.269)	(0.071)	(0.921)	(0.959)	(0.079)
Issue Amount	6.319***	-16.56***	8.237***	-46.70***	9.010***	5.338***
	(0.978)	(5.432)	(0.816)	(12.84)	(2.147)	(1.178)
Sample	Low Cost	High Cost	IG	HY	ST	LT
Issuer-Day FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.461	0.433	0.331	0.505	0.465	0.368
Observations	3,303,875	563,984	3,527,369	835,465	485,363	3,381,645

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table reports the 2SLS regression estimates. Panel A reports the estimates from the first-stage regressions of transaction costs (measured in basis points) on the instrument $New\ Bond_{it}$. Panel B reports the estimates from the second-stage regressions of credit spreads on the instrumented transaction costs. The instrumented transaction costs are the predicted values obtained from the first-stage regressions. Observations are at the bond-day level. All specifications control for bond characteristics including time-to-maturity measured in years and the logarithm of issue amount, and include the issuer-day fixed effects. Standard errors, which are clustered at the bond and the day levels, are reported in parentheses. The estimation sample is constructed using the enhanced TRACE data and the Mergent FISD data from 1/1/2005 to 12/31/2021. Sub-samples include bonds whose transaction costs in the prior month are below the 75th percentile (Low Cost), bonds whose transaction costs in the prior month are above the 75th percentile (High Cost), bonds rated investment-grade (IG), bonds rated high-yield (HY), short-term bonds with time-to-maturity below 3 years (ST), and long-term bonds with time-to-maturity above 3 years (LT).

maturity date of the given new bond. In line with the previous results, the average credit spread differential between the more liquid new bonds and the less liquid matched old bonds is mostly positive throughout the sample period. However, and consistent with Prediction 1b, the credit spread differential is noticeably negative from Q4 2008 to Q2 2009, following the Lehman Brothers collapse and in the depth of the Great Financial Crisis. The differential remains positive in Q1 2020 during the COVID-19 shock. This may be due to the Fed's swift responses, such as the announcement of the Corporate Credit Facilities, which quickly reversed the credit market conditions.[20]

Figure 5: Credit Spread Differential Between Matched Old and New Bonds (in bps)



Note: This figure plots the quarterly time series of the average credit spread differential between new bonds and matched old bonds from Q1 2006 to Q4 2021. New bonds are bonds with age below one year. Olde bonds are the same issuers' old bonds aged three years or above, and with maturity dates within one year of that of the new bond. Corporate bond transaction data are from the enhanced TRACE and bond characteristics data are from the Mergent FISD.

5.3 Asymmetric Transaction Costs

In this section, I compare the transaction costs incurred by selling investors and those incurred by buying investors. I compute the average transaction costs at the bond-trade size-trade direction-day level.[21] As before, I follow the literature and exclude observations with time-to-maturity

[20] E.g., Gilchrist et al. (2021), O'Hara and Zhou (2021), Boyarchenko et al. (2022).

[21] I follow the literature convention as well as industry practice and divide trade sizes into four categories, including micro trades (<\$100,000), odd-lot trades (\$100,000 - \$1 million), round-lot trades (\$1 - 5 million) and block trades ($\geq$ \$5 million). Micro trades are considered retail-sized, and are excluded from my analysis.

below one year and observations with missing credit ratings. To minimize noise, I also exclude observations with defaulted bonds (that is, bonds rated D or below) from the sample.

To estimate the transaction cost differential between investor-buy trades and investor-sell trades, I estimate the following specification

$$Transaction\,Cost_{itds} = \delta Investor\ Sell_{itds} + \alpha_{its} + \epsilon_{itds} \tag{5.4}$$

where i indexes bond, t indexes day, d indexes trade direction, and s indexes trade size. $Transaction\,Cost_{itds}$ is the average transaction cost of bond i with trade direction d and trade size s on day t. $Investor\ Sell_{itds}$ is a dummy variable indicating the direction of the trade, and equals one if the trade is an investor sale and zero if the trade is an investor purchase. The key controls are the issue-trade size-day fixed effects α_{its}, which control for time-varying impact of bond-level and trade-level characteristics on transaction costs. The specification above thus compares investor-buy and investor-sell trades of the same size bucket in the same bond on the same day. Standard errors are clustered at the bond and the day levels.

The coefficient of interest is δ. If selling investors incur higher transaction costs than buying investors, then δ should be positive. The regression results are presented in Table 4. Column (1) reports the coefficient estimate for the full sample. The coefficient estimate is indeed positive, and statistically significant at the 10% level. The result suggests that the transaction costs of investor-sell trades are 2.5 bps higher on average than those of investor-buy trades. Given that the average transaction cost is just over 20 bps, the transaction cost differential of 2.5 bps is substantial in magnitude.

The full sample includes periods when market uncertainty is low and investors are mostly buying. During these periods, selling investors may not be significantly more impatient than buying investors. Moreover, factors outside the model suggest that investor-sell trades may incur lower transaction costs during these specific periods. For instance, high buying pressure by investors leads to negative dealer inventories in the short run, and dealers tend to charge lower transaction costs for investor-sell trades that offset their inventory imbalances. Hence, I exclude

Table 4: Transaction Cost Asymmetry Between Investor-Sell and Investor-Buy Trades

	Transaction Cost (bps)			
	(1)	(2)	(3)	(4)
Investor Sell	2.544*	6.161***	4.523***	3.337**
	(1.376)	(1.936)	(1.701)	(1.537)
Sample	Full	VIX Filter	DEF Filter	B/S Filter
Issue-Day-Size FE	Yes	Yes	Yes	Yes
R^2	0.050	0.044	0.038	0.052
Observations	6,591,026	4,556,170	5,203,854	5,374,874

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table reports the regression estimates examining the differential transaction costs between investor-buy trades and investor-sell trades. *Investor Sell* is a dummy variable equal to one if the trade is an investor sale and zero otherwise. Observations are at the bond-trade size-trade direction-day level. All regressions control for the issue-trade size-day fixed effects. Standard errors, which are clustered at the bond and the day levels, are reported in parentheses. Transaction data are from the enhanced TRACE. The full estimation sample covers trades from 1/1/2005 to 12/31/2021. VIX Filter excludes periods when the VIX value is below the 25th percentile. DEF Filter excludes periods when the credit spread differential between BBB-rated corporate bonds and AAA-rated corporate bonds is below the 25th percentile. B/S Filter excludes periods when the investor buy-sell ratio is above the 75th percentile.

these periods from the sample and re-estimate the specification (5.4). The results are provided in columns (2)-(4) of Table 4. Column (2) excludes periods when the VIX index value is below the 25th percentile (around 13.96) from the estimation sample. Column (3) excludes periods when the credit spread difference between BBB-rated corporate bonds and AAA-rated corporate bonds is below the 25th percentile (around 79 bps) from the estimation sample. Column (4) excludes periods when the investor buy-sell ratio is above the 75th percentile (around 1.65) from the sample. In all columns, the coefficient estimates are positive and highly statistically significant. On average, selling investors incur 3-6 bps higher transaction costs than buying investors.

6 Conclusion

Many assets, such as fixed income securities, repurchase agreements and derivatives, are traded over-the-counter in intermediated markets, which feature trading illiquidity due to imperfect search and matching as well as strategic bargaining between investors and dealers. At the same time, these assets are often systemically important, due to their size and significance in financing to firms and financial institutions. In this paper, I examine the effect of asset liquidity on portfolio choice and asset prices in intermediated markets. To do so, I build a highly tractable search-theoretic model that accommodates both strategic bargaining and portfolio choice. Tractability stems from investors' quadratic preferences, the semi-centralized market structure, and i.i.d. asset cash flows. The model generates several predictions novel to the literature. The model predicts that the relationship between asset prices and asset liquidity in intermediated markets is non-monotonic. The price-liquidity relationship is positive for sufficiently liquid assets, but negative for highly illiquid assets. The average price-liquidity relation may turn negative during severe crises. The model also predicts that selling investors incur higher transaction costs than buying investors.

The non-monotonic price-liquidity relationship, in particular, contrasts with conventional wisdom. Strategic bargaining, a feature of the intermediated markets, plays a crucial role in generating the non-monotonicity. Hence, the paper sheds light on how market structure can affect not only short-term market outcomes but also long-term equilibrium asset prices. Thus, policies aimed at boosting market liquidity should take into account market structure and design issues. More research is needed, however, to understand how specific market structure features interact with (and potentially amplify) other financial frictions in affecting equilibrium outcomes.

References

Acharya, Viral V. and Lasse Heje Pedersen, "Asset pricing with liquidity risk," *Journal of Financial Economics*, 2005, *77* (2), 375–410.

_ , Yakov Amihud, and Sreedhar T. Bharath, "Liquidity risk of corporate bond returns: conditional approach," *Journal of Financial Economics*, 2013, *110* (2), 358–386.

Afonso, Gara and Ricardo Lagos, "Trade Dynamics in the Market for Federal Funds," *Econometrica*, 2015, *83* (1), 263–313.

Aiyagari, S. Rao and Mark Gertler, "Asset Returns with Transaction Costs and Uninsured Individual Risk," *Journal of Monetary Economics*, 1991, *27* (3), 311–331.

Amihud, Yakov and Haim Mendelson, "Asset Pricing and the Bid-Ask Spread," *Journal of Financial Economics*, 1986, *17* (2), 223–249.

_ and _ , "Liquidity, Maturity, and the Yields on U.S. Treasury Securities," *Journal of Finance*, 1991, *46* (4), 1411–1425.

An, Yu, "Competing with Inventory in Dealership Markets," Working Paper 2022.

Atkeson, Andrew G., Andrea L. Eisfeldt, and Pierre-Olivier Weill, "Entry and Exit in OTC Derivatives Markets," *Econometrica*, 2015, *83* (6), 2231–2292.

Boudoukh, Jacob, Jordan Brooks, Matthew Richardson, and Xu Zhikai, "Sovereign credit quality and violations of the law of one price," Working Paper 2019.

Boyarchenko, Nina, Anna Kovner, and Or Shachar, "It's what you say and what you buy: A holistic evaluation of the corporate credit facilities," *Journal of Financial Economics*, 2022, *144* (3), 695–731.

Chen, Long, David A. Lesmond, and Jason Wei, "Corporate Yield Spreads and Bond Liquidity," *Journal of Finance*, 2007, *62* (1), 119–149.

Choi, Jaewon, Jungsuk Han, Sean S. Shin, and Ji Hee Yoon, "The More Illiquid, the More Expensive: A Search-Based Explanation of the Illiquidity Premium," Working Paper 2022.

___, **Or Shachar, and Sean Seunghun Shin**, "Dealer Liquidity Provision and the Breakdown of the Law of One Price: Evidence from the CDS-Bond Basis," *Management Science*, 2019, *65* (9), 4100–4122.

Constantinides, George M., "Capital Market Equilibrium with Transaction Costs," *Journal of Political Economy*, 1986, *94* (4), 842–862.

Dick-Nielsen, Jens, "How to clean Enhanced TRACE data," Working Paper 2014.

Duffie, Darrell, Nicolae Garleanu, and Lasse Heje Pedersen, "Over-the-Counter Markets," *Econometrica*, 2005, *73* (6), 1815–1847.

___, ___, **and** ___, "Valuation in Over-the-Counter Markets," *Review of Financial Studies*, 2007, *20* (6), 1865–1900.

Edwards, Amy K., Lawrence E. Harris, and Michael S. Piwowar, "Corporate Bond Market Transaction Costs and Transparency," *Journal of Finance*, 2007, *62* (3), 1421–1451.

Farboodi, Maryam, Gregor Jarosch, Guido Menzio, and Ursula Wiriadinata, "Intermediation as Rent Extraction," Working Paper 2019.

Favero, Carlo, Marco Pagano, and Ernst-Ludwig von Thadden, "How Does Liquidity Affect Government Bond Yields?," *Journal of Financial and Quantitative Analysis*, 2010, *45* (1), 107–134.

Garleanu, Nicolae, "Portfolio Choice and Pricing in Illiquid Markets," *Journal of Economic Theory*, 2009, *144* (2), 532–564.

Gilchrist, Simon, Bin Wei, Vivian Z. Sue, and Egon Zakrajšek, "The Fed Takes on Corporate Credit Risk: An Analysis of the Efficacy of the SMCCF," Working Paper 2021.

Gürkaynak, Refet S., Brian Sack, and Jonathan H. Wright, "The U.S. Treasury Yield Curve: 1961 to the Present," *Journal of Monetary Economics*, 2007, *54* (8), 2291–2304.

Heaton, John and Deborah J. Lucas, "Evaluating the Effects of Incomplete Markets on Risk Sharing and Asset Pricing," *Journal of Political Economy*, 1996, *104* (3), 443–487.

Hendershott, Terrence and Ananth Madhavan, "Click or Call? Auction versus Search in the Over-the-Counter Market," *Journal of Finance*, 2015, *70* (1), 419–447.

Houweling, Patrick, Albert Mentink, and Ton Vorst, "Comparing possible proxies of corporate bond liquidity," *Journal of Banking & Finance*, 2005, *29* (6), 1331–1358.

Hugonnier, Julien, Benjamin Lester, and Pierre-Olivier Weill, "Frictional Intermediation in Over-the-Counter Markets," *Review of Economic Studies*, 2020, *87* (3), 1432–1469.

Lagos, Ricardo and Guillaume Rocheteau, "Liquidity in Asset Markets with Search Frictions," *Econometrica*, 2009, *77* (2), 403–426.

Lin, Hai, Junbo Wang, and Chunchi Wu, "Liquidity risk and expected corporate bond returns," *Journal of Financial Economics*, 2011, *99* (3), 628–650.

Longstaff, Francis A., Sanjay Mithal, and Eric Neis, "Corporate Yield Spreads: Default Risk or Liquidity? New Evidence from the Credit Default Swap Market," *Journal of Finance*, 2005, *60* (5), 2213–2252.

Ma, Yiming, Kairong Xiao, and Yao Zeng, "Mutual Fund Liquidity Transformation and Reverse Flight to Liquidity," *Review of Financial Studies*, 2022, *35* (10), 4674–4711.

Milbradt, Konstantin, "Asset Heterogeneity in Over-the-Counter Markets," Working Paper 2017.

Mizrach, Bruce, "Analysis of Corporate Bond Liquidity," FINRA Office of the Chief Economist Research Note 2015.

O'Hara, Maureen and Xing (Alex) Zhou, "Anatomy of a liquidity crisis: Corporate bonds in the COVID-19 crisis," *Journal of Financial Economics*, 2021, *142* (1), 46–68.

Pasquariello, Paolo and Clara Vega, "The on-the-run liquidity phenomenon," *Journal of Financial Economics*, 2009, *92* (1), 1–24.

Pastor, Lubos and Robert F. Stambaugh, "Liquidity Risk and Expected Stock Returns," *Journal of Political Economy*, 2003, *111* (3), 642–685.

Rubinstein, Ariel, "Perfect Equilibrium in a Bargaining Model," *Econometrica*, 1982, *50* (1), 97–109.

_ **and Asher Wolinsky**, "Equilibrium in a Market with Sequential Bargaining," *Econometrica*, 1985, *53* (5), 1133–1150.

Sambalaibat, Batchimeg, "A Theory of Liquidity Spillover between Bond and CDS Markets," *Review of Financial Studies*, 2022, *35* (5), 2525–2569.

Sarig, Oded and Arthur Warga, "Bond Price Data and Bond Market Liquidity," *Journal of Financial and Quantitative Analysis*, 1989, *24* (3), 367–378.

Üslü, Semih, "Pricing and Liquidity in Decentralized Asset Markets," *Econometrica*, 2019, *87* (6), 2079–2140.

_ **and Güner Velioğlu**, "Liquidity in the Cross Section of OTC Assets," Working Paper 2021.

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

_ **and Pierre-Olivier Weill**, "A Search-Based Theory of the On-the-Run Phenomenon," *Journal of Finance*, 2008, *63* (3), 1361–1398.

_ **and Tan Wang**, "Search and Endogenous Concentration of Liquidity in Asset Markets," *Journal of Economic Theory*, 2007, *136* (1), 66–104.

Weill, Pierre-Olivier, "Liquidity Premia in Dynamic Bargaining Markets," *Journal of Economic Theory*, 2008, *140* (1), 66–96.

Appendix

A.1 Strategic Bargaining Game

The bargaining game occurs in virtual time. Let d_t be the time interval between bargaining rounds. I first treat d_t as a small discrete increment, and then return to the continuous time case by setting $d_t \to 0$. Consider an investor with patience type $\xi \in \{h, l\}$ and asset holdings $\mathbf{x}$, who meets a dealer to trade asset i. Let $V_\xi(\mathbf{x})$ denote the investor's expected value. The investor discounts time at rate $\beta \equiv r + \epsilon \mathbb{I}_{\{\xi=l\}}$, where $\mathbb{I}_{\{\xi=l\}}$ equals one if the investor is type l and zero otherwise. Let $P^i = P^i_\xi(\mathbf{x})$ and $q^i = q^i_\xi(\mathbf{x})$ denote the trade price and quantity. The investor's gain from trade is

$$V_\xi(x^1, \dots, x^i + q^i, \dots, x^N) - V_\xi(\mathbf{x}) - P^i q^i \tag{A.1}$$

$\mathcal{G}_\xi(P^i, q^i|\mathbf{x}) \equiv V_\xi(x^1, \dots, x^i + q^i, \dots, x^N) - V_\xi(\mathbf{x})$ is the change in the investor's value due to purchasing q^i units of the asset. $P^i q^i$ is the amount paid by the investor. The dealer's gain from trade is its intermediation profit from trading asset i, that is

$$P^i q^i - \bar{P}^i q^i \tag{A.2}$$

where $\bar{P}^i$ is the inter-dealer market price of the asset.

When a party is chosen to make an offer, the selected party suggests both a trade price and a trade quantity. Let P^i_c and q^i_c denote the price and quantity suggested by the investor (or customer). Let P^i_d and q^i_d denote the price and quantity suggested by the dealer. Given the dealer's and the investor's offer probabilities, the expected price, quantity and trade amount (product of price and quantity) are

$$\mathbb{E}[P^i] = (1-z)P^i_c + zP^i_d \tag{A.3}$$

$$\mathbb{E}[q^i] = (1-z)q^i_c + zq^i_d \tag{A.4}$$

$$\mathbb{E}[P^i q^i] = (1-z)P^i_c q^i_c + zP^i_d q^i_d \tag{A.5}$$

Similarly, the investor's expected change in value due to trading is $\mathbb{E}[\mathcal{G}_\xi(P^i, q^i|\mathbf{x})] = (1-z)\mathcal{G}_\xi(P^i_c, q^i_c|\mathbf{x}) + z\mathcal{G}_\xi(P^i_d, q^i_d|\mathbf{x})$. The selected party makes an offer that leaves the other party indifferent between accepting and rejecting the offer. That is,

$$\mathcal{G}_\xi(P_d^i, q_d^i|\mathbf{x}) - P_d^i q_d^i = e^{-\beta d_t}(\mathbb{E}[\mathcal{G}_\xi(P^i, q^i|\mathbf{x})] - \mathbb{E}[P^i q^i]) + O(d_t^2) \tag{A.6}$$

$$P_c^i q_c^i - \bar{P}^i q_c^i = e^{-r d_t}(\mathbb{E}[P^i q^i] - \bar{P}^i \mathbb{E}[q^i]) + O(d_t^2) \tag{A.7}$$

For example, in (A.6), the dealer is selected to make an offer and suggests $\{P_d^i, q_d^i\}$. If the investor accepts the offer, their immediate gain from trade is $\mathcal{G}_\xi(P_d^i, q_d^i|\mathbf{x}) - P_d^i q_d^i$. If the investor rejects the offer, their expected gain from trade in the next round is $\mathbb{E}[\mathcal{G}_\xi(P^i, q^i|\mathbf{x}) - P^i q^i]$. The present value of rejecting the offer is the expected gain from trade discounted at the investor's discount rate β. Setting $d_t \to 0$, in the limit, $\lim_{d_t \to 0} P_c^i = \lim_{d_t \to 0} P_d^i \equiv P^i$ and $\lim_{d_t \to 0} q_c^i = \lim_{d_t \to 0} q_d^i \equiv q^i$. Thus, $\lim_{d_t \to 0} P_c^i q_c^i = \lim_{d_t \to 0} P_d^i q_d^i = P^i q^i$, and $\lim_{d_t \to 0} \mathcal{G}_\xi(P_c^i, q_c^i|\mathbf{x}) = \lim_{d_t \to 0} \mathcal{G}_\xi(P_d^i q_d^i|\mathbf{x}) = \mathcal{G}_\xi(P^i, q^i|\mathbf{x})$ since $\mathcal{G}_\xi(\cdot, \cdot|\mathbf{x})$ is continuous. From (A.6) and (A.7),

$$\begin{aligned} P^i q^i &= \frac{z(1 - e^{-\beta d_t})\mathcal{G}_\xi(P^i, q^i|\mathbf{x}) + (1 - z)(1 - e^{-r d_t})\bar{P}^i q^i}{z(1 - e^{-\beta d_t}) + (1 - z)(1 - e^{-r d_t})} \\ &= \frac{z\beta}{z\beta + (1 - z)r}\mathcal{G}_\xi(P^i, q^i|\mathbf{x}) + \frac{(1 - z)r}{z\beta + (1 - z)r}\bar{P}^i q^i \end{aligned} \tag{A.8}$$

The second equality follows from $\lim_{d_t \to 0} 1 - e^{-y d_t} = y d_t$. Let

$$\theta_\xi \equiv \frac{(1 - z)r}{(1 - z)r + z\beta} = \frac{(1 - z)r}{r + z\epsilon \mathbb{I}_{\{\xi = l\}}} \tag{A.9}$$

It is immediate that $\theta_h > \theta_l$. Then, (A.8) can be written as

$$P^i q^i = (1 - \theta_\xi)\Big[V_\xi(x^1, \ldots, x^i + q^i, \ldots, x^N) - V_\xi(\mathbf{x})\Big] + \theta_\xi \bar{P}^i q^i \tag{A.10}$$

Substituting (A.10) into (A.1) yields the investor's gain from trade $\theta_\xi\Big[V_\xi(x^1, \ldots, x^i + q^i, \ldots, x^N) - V_\xi(\mathbf{x}) - \bar{P}^i q^i\Big]$. Substituting (A.10) into (A.2) gives the dealer's gain from trade $(1 - \theta_\xi)\Big[V_\xi(x^1, \ldots, x^i + q^i, \ldots, x^N) - V_\xi(\mathbf{x}) - \bar{P}^i q^i\Big]$. Pareto optimality requires that the quantity maximizes $V_\xi(x^1, \ldots, x^i + q^i, \ldots, x^N) - V_\xi(\mathbf{x}) - \bar{P}^i q^i$, and thus

$$V_{\xi i}(x^1, \ldots, x^i + q^i, \ldots, x^N) = \bar{P}^i \tag{A.11}$$

where the subscript i denotes derivative with respect to the i-th argument. If the selected party suggests a quantity other than what is given above, the party can make both parties better off by choosing q as given by (A.11) instead.

Following Rubinstein (1982), this is a unique subgame perfect equilibrium.

A.2 Proof of Proposition 2

I conjecture (and verify) that $V_h(x)$ is quadratic in its arguments, that is,

$$V_h(\mathbf{x}) = H_h + \sum_i I_h^i x^i + \sum_i L_h^i (x^i)^2 + \sum_{j \neq i} L_h^{ij} (x^i)(x^j) \tag{A.12}$$

Differentiate with respect to x^i,

$$V_{hi}(\mathbf{x}) = I_h^i + 2L_h^i x^i + \sum_{j \neq i} (L_h^{ij} + L_h^{ji}) x^j \tag{A.13}$$

Note that for any q^i and $i \in \{1, 2, \ldots, N\}$,

$$\begin{aligned} &V_h\big(x^1, \ldots, x^i{+}q^i, \ldots, x^N\big) - V_h(\mathbf{x}) \\ &\qquad = V_{hi}\big(x^1, \ldots, x^i + q^i, \ldots, x^N\big) q^i - L_h^i (q^i)^2 \end{aligned} \tag{A.14}$$

Moreover, from (3.4),

$$V_{hi}\big(x^1, \ldots, x^i + q^i, \ldots, x^N\big) = \bar{P}^i \tag{A.15}$$

Substituting (A.13) into the left-hand side of (A.15) and rearranging gives

$$q_h^i(\mathbf{x}) = \frac{\bar{P}^i - V_{hi}(\mathbf{x})}{2L_h^i} \tag{A.16}$$

The joint trade surplus between the type h investor and the dealer becomes

$$\begin{aligned} &V_h\big(x^1, \ldots, x^i{+}q_h^i(\mathbf{x}), \ldots, x^N\big) - V_h(\mathbf{x}) - \bar{P}^i q_h^i(\mathbf{x}) \\ &\qquad = -L_h^i q_h^i(\mathbf{x})^2 = -\frac{[\bar{P}^i - V_{hi}(\mathbf{x})]^2}{4L_h^i} \end{aligned} \tag{A.17}$$

The first equality follows from (A.14) and (A.15). The HJB equation (3.9) becomes

$$\begin{aligned} rV_h(\mathbf{x}) = \sum_{i=1}^{N} \bar{D}x^i - \frac{\gamma}{2}\sum_{i=1}^{N} \sigma^2 (x^i)^2 + \zeta_{hl}\Big[V_l(\mathbf{x}) - V_h(\mathbf{x})\Big] \\ + \sum_{i=1}^{N} \lambda^i \theta_h \Big[-\frac{[\bar{P}^i - V_{hi}(\mathbf{x})]^2}{4L_h^i}\Big] \end{aligned} \tag{A.18}$$

This is the HJB equation for a type h investor with asset holdings $\mathbf{x}$. I then differentiate with respect to x^i on both sides to get the HJB equation for the marginal value of the investor,

$$\begin{aligned} rV_{hi}(\mathbf{x}) =& \bar{D} - \gamma\sigma^2 x^i + \zeta_{hl}\Big[V_{li}(\mathbf{x}) - V_{hi}(\mathbf{x})\Big] + \lambda^i\theta_h\Big[\bar{P}^i - V_{hi}(\mathbf{x})\Big] \\ &+ \sum_{j\neq i} \lambda^j \theta_h \frac{L_h^{ij} + L_h^{ji}}{2L_h^j}\Big[\bar{P}^j - V_{hj}(\mathbf{x})\Big] \end{aligned} \tag{A.19}$$

I similarly conjecture (and verify) that $V_l(\mathbf{x})$ is quadratic in its arguments, that is,

$$V_l(\mathbf{x}) = H_l + \sum_i I_l^i x^i + \sum_i L_l^i (x^i)^2 + \sum_{j\neq i} L_l^{ij}(x^i)(x^j) \tag{A.20}$$

Take derivative with respect to x^i,

$$V_{li}(\mathbf{x}) = I_l^i + 2L_l^i x^i + \sum_{j\neq i}(L_l^{ij} + L_l^{ji})x^j \tag{A.21}$$

I follow the derivations above to rewrite the HJB equation (3.10) and then differentiate with respect to x^i on both sides to get the HJB equation for the marginal value of the type l investor with asset holdings $\mathbf{x}$,

$$\begin{aligned} (r+\epsilon)V_{li}(\mathbf{x}) =& \bar{D} - \gamma\sigma^2 x^i + \zeta_{lh}\Big[V_{hi}(\mathbf{x}) - V_{li}(\mathbf{x})\Big] + \lambda^i\theta_l\Big[\bar{P}^i - V_{li}(\mathbf{x})\Big] \\ &+ \sum_{j\neq i} \lambda^j \theta_l \frac{L_l^{ij} + L_l^{ji}}{2L_l^j}\Big[\bar{P}^j - V_{lj}(\mathbf{x})\Big] \end{aligned} \tag{A.22}$$

Substitute (A.13) and (A.21) into (A.19),

$$
\begin{aligned}
(r+\zeta_{hl}+\lambda^i\theta_h)\Big[I_h^i+2L_h^i x^i+\sum_{j\neq i}(L_h^{ij}+L_h^{ji})x^j\Big] &= \bar{D}-\gamma\sigma^2 x^i \\
+\zeta_{hl}\Big[I_l^i+2L_l^i x^i+\sum_{j\neq i}(L_l^{ij}+L_l^{ji})x^j\Big] &+ \Big[\lambda^i\theta_h\bar{P}^i+\sum_{j\neq i}\lambda^j\theta_h\frac{L_h^{ij}+L_h^{ji}}{2L_h^j}\bar{P}^j\Big] \\
-\sum_{j\neq i}\lambda^j\theta_h\frac{L_h^{ij}+L_h^{ji}}{2L_h^i}\Big[I_h^j+2L_h^j x^j &+\sum_{k\neq j}(L_h^{jk}+L_h^{kj})x^k\Big]
\end{aligned}
\tag{A.23}
$$

Substitute (A.13) and (A.21) into (A.22),

$$
\begin{aligned}
(r+\epsilon+\zeta_{lh}+\lambda^i\theta_l)\Big[I_l^i+2L_l^i x^i+\sum_{j\neq i}(L_l^{ij}+L_l^{ji})x^j\Big] &= \bar{D}-\gamma\sigma^2 x^i \\
+\zeta_{lh}\Big[I_h^i+2L_h^i x^i+\sum_{j\neq i}(L_h^{ij}+L_h^{ji})x^j\Big] &+ \Big[\lambda^i\theta_l\bar{P}^i+\sum_{j\neq i}\lambda^j\theta_l\frac{L_l^{ij}+L_l^{ji}}{2L_l^j}\bar{P}^j\Big] \\
-\sum_{j\neq i}\lambda^j\theta_l\frac{L_l^{ij}+L_l^{ji}}{2L_l^j}\Big[I_l^j+2L_l^j x^j &+\sum_{k\neq j}(L_l^{jk}+L_l^{kj})x^k\Big]
\end{aligned}
\tag{A.24}
$$

From the two equations above, and use the method of undetermined coefficients, $\forall\, j\neq i\in\{1,2,\ldots,N\}$,

$$\Big[r+\zeta_{hl}+(\lambda^i+\lambda^j)\theta_h\Big](L_h^{ij}+L_h^{ji})=\zeta_{hl}(L_l^{ij}+L_l^{ji}) \tag{A.25}$$

$$\Big[r+\epsilon+\zeta_{lh}+(\lambda^i+\lambda^j)\theta_l\Big](L_l^{ij}+L_l^{ji})=\zeta_{lh}(L_h^{ij}+L_h^{ji}) \tag{A.26}$$

Hence, it must be that

$$L_h^{ij}+L_h^{ji}=L_l^{ij}+L_l^{ji}=0 \qquad \forall\, j\neq i \tag{A.27}$$

Substitute into (A.19) and (A.22), and rewrite the two equations as

$$rV_{hi}(\mathbf{x})=\bar{D}-\gamma\sigma^2x^i+\zeta_{hl}\Big[V_{li}(\mathbf{x})-V_{hi}(\mathbf{x})\Big]+\lambda^i\theta_h\Big[\bar{P}^i-V_{hi}(\mathbf{x})\Big] \tag{A.28}$$

$$(r+\epsilon)V_{li}(\mathbf{x})=\bar{D}-\gamma\sigma^2x^i+\zeta_{lh}\Big[V_{hi}(\mathbf{x})-V_{li}(\mathbf{x})\Big]+\lambda^i\theta_l\Big[\bar{P}^i-V_{li}(\mathbf{x})\Big] \tag{A.29}$$

Rearranging (A.28) and (A.29) yields

$$(r+\zeta_{hl}+\lambda^i\theta_h)V_{hi}(\mathbf{x})-\zeta_{hl}V_{li}(\mathbf{x})=\bar{D}-\gamma\sigma^2x^i+\lambda^i\theta_h\bar{P}^i \tag{A.30}$$

$$(r+\epsilon+\zeta_{lh}+\lambda^i\theta_l)V_{li}(\mathbf{x})-\zeta_{lh}V_{hi}(\mathbf{x})=\bar{D}-\gamma\sigma^2x^i+\lambda^i\theta_l\bar{P}^i \tag{A.31}$$

Solve for $V_{hi}(\mathbf{x})$ and $V_{li}(\mathbf{x})$,

$$\begin{aligned}V_{hi}(\mathbf{x})=&\frac{(r+\epsilon+\zeta_{lh}+\zeta_{hl}+\lambda^i\theta_l)(\bar{D}-\gamma\sigma^2x^i)}{(r+\zeta_{hl}+\lambda^i\theta_h)(r+\epsilon+\zeta_{lh}+\lambda^i\theta_l)-\zeta_{hl}\zeta_{lh}}\\&+\frac{(r+\epsilon+\zeta_{lh}+\lambda^i\theta_l)\lambda^i\theta_h+\zeta_{hl}\lambda^i\theta_l}{(r+\zeta_{hl}+\lambda^i\theta_h)(r+\epsilon+\zeta_{lh}+\lambda^i\theta_l)-\zeta_{hl}\zeta_{lh}}\bar{P}^i\end{aligned} \tag{A.32}$$

and

$$\begin{aligned}V_{li}(\mathbf{x})=&\frac{(r+\zeta_{lh}+\zeta_{hl}+\lambda^i\theta_h)(\bar{D}-\gamma\sigma^2x^i)}{(r+\zeta_{hl}+\lambda^i\theta_h)(r+\epsilon+\zeta_{lh}+\lambda^i\theta_l)-\zeta_{hl}\zeta_{lh}}\\&+\frac{\zeta_{lh}\lambda^i\theta_h+(r+\zeta_{hl}+\lambda^i\theta_h)\lambda^i\theta_l}{(r+\zeta_{hl}+\lambda^i\theta_h)(r+\epsilon+\zeta_{lh}+\lambda^i\theta_l)-\zeta_{hl}\zeta_{lh}}\bar{P}^i\end{aligned} \tag{A.33}$$

By the method of undetermined coefficients,

$$L_h^i=-\frac{1}{2}\frac{(r+\epsilon+\zeta_{lh}+\zeta_{hl}+\lambda^i\theta_l)\gamma\sigma^2}{(r+\zeta_{hl}+\lambda^i\theta_h)(r+\epsilon+\zeta_{lh}+\lambda^i\theta_l)-\zeta_{hl}\zeta_{lh}} \tag{A.34}$$

$$L_l^i=-\frac{1}{2}\frac{(r+\zeta_{lh}+\zeta_{hl}+\lambda^i\theta_h)\gamma\sigma^2}{(r+\zeta_{hl}+\lambda^i\theta_h)(r+\epsilon+\zeta_{lh}+\lambda^i\theta_l)-\zeta_{hl}\zeta_{lh}} \tag{A.35}$$

Subtract the inter-dealer price $\bar{P}^i$ from (A.32) and (A.33),

$$V_{hi}(\mathbf{x})-\bar{P}^i=\frac{(r+\epsilon+\zeta_{lh}+\zeta_{hl}+\lambda^i\theta_l)(\bar{D}-\gamma\sigma^2x^i-r\bar{P}^i)-\epsilon\zeta_{hl}\bar{P}^i}{(r+\zeta_{hl}+\lambda^i\theta_h)(r+\epsilon+\zeta_{lh}+\lambda^i\theta_l)-\zeta_{hl}\zeta_{lh}} \tag{A.36}$$

$$V_{li}(\mathbf{x})-\bar{P}^i=\frac{(r+\zeta_{lh}+\zeta_{hl}+\lambda^i\theta_h)[\bar{D}-\gamma\sigma^2x^i-(r+\epsilon)\bar{P}^i]+\epsilon\zeta_{lh}\bar{P}^i}{(r+\zeta_{hl}+\lambda^i\theta_h)(r+\epsilon+\zeta_{lh}+\lambda^i\theta_l)-\zeta_{hl}\zeta_{lh}} \tag{A.37}$$

Let $x_h^i(\mathbf{x})$ and $x_l^i(\mathbf{x})$ be the optimal asset i demand by a type h investor with asset holdings $\mathbf{x}$ and a type l investor with asset holdings $\mathbf{x}$ respectively. $V_{hi}(\mathbf{x})|_{x^i=x_h^i(\mathbf{x})}=\bar{P}^i$ and $V_{li}(\mathbf{x})|_{x^i=x_l^i(\mathbf{x})}=\bar{P}^i$. Solving for $x_h^i(\mathbf{x})$ and $x_l^i(\mathbf{x})$, I get

$$x_h^i(\mathbf{x}) \equiv x_h^i = \frac{\bar{D} - (r + \frac{\zeta_{hl}\epsilon}{r+\epsilon+\zeta_{lh}+\zeta_{hl}+\lambda^i\theta_l})\bar{P}^i}{\gamma\sigma^2} \tag{A.38}$$

$$x_l^i(\mathbf{x}) \equiv x_l^i = \frac{\bar{D} - (r + \epsilon - \frac{\zeta_{lh}\epsilon}{r+\zeta_{lh}+\zeta_{hl}+\lambda^i\theta_h})\bar{P}^i}{\gamma\sigma^2} \tag{A.39}$$

A.3 Proof of Proposition 3

The steady-state distribution $\Phi(\mathbf{x}, \xi)$ is unique and determined by (3.15) and the fact that investors' measures sum to one. For any asset i, the optimal holdings are given by (A.38) and (A.39), which are strictly decreasing functions of the inter-dealer price of the asset $\bar{P}^i$. The market clearing condition (3.16) thus uniquely determines the price $\bar{P}^i$. For any asset i, given $\bar{P}^i$, x_h^i and x_l^i that satisfy (A.38) and (A.39) are also uniquely determined. Given $\bar{P}^i$, x_h^i and x_l^i, for all $i \in \{1, 2, \ldots, N\}$, the trade prices $P_h^i(\mathbf{x})$ and $P_l^i(\mathbf{x})$ and trade quantities $q_h^i(\mathbf{x})$ and $q_l^i(\mathbf{x})$ are also uniquely determined. Therefore, there exists a unique stationary equilibrium in the economy.

A.4 Proof of Proposition 5

Combine (3.17)-(3.20) and (3.21),

$$\mu(x_h^i, h) + \mu(x_h^i, l) + \mu(x_l^i, h) + \mu(x_l^i, l) = 1 \tag{A.40}$$

$$-\zeta_{hl}\mu(x_h^i, h) + \zeta_{lh}\mu(x_h^i, l) + \lambda^i\mu(x_l^i, h) = 0 \tag{A.41}$$

$$-\zeta_{hl}\mu(x_l^i, h) + \zeta_{lh}\mu(x_l^i, l) - \lambda^i\mu(x_l^i, h) = 0 \tag{A.42}$$

$$-\zeta_{lh}\mu(x_h^i, l) + \zeta_{hl}\mu(x_h^i, h) - \lambda^i\mu(x_h^i, l) = 0 \tag{A.43}$$

$$-\zeta_{lh}\mu(x_l^i, l) + \zeta_{hl}\mu(x_l^i, h) + \lambda^i\mu(x_h^i, l) = 0 \tag{A.44}$$

Notice that one of the equations is redundant. Thus, there is a unique solution to the system of linear equations. Add up (A.41) and (A.42), using (A.40),

$$\mu(x_h^i, h) + \mu(x_l^i, h) = \frac{\zeta_{lh}}{\zeta_{hl} + \zeta_{lh}} = 1 - \pi \tag{A.45}$$

$$\mu(x_h^i, l) + \mu(x_l^i, l) = \frac{\zeta_{hl}}{\zeta_{hl} + \zeta_{lh}} = \pi \tag{A.46}$$

Rearrange (A.42) and (A.43),

$$(\zeta_{hl} + \lambda^i)\mu(x_l^i, h) = \zeta_{lh}\mu(x_l^i, l) = \zeta_{lh}(\pi - \mu(x_h^i, l)) \tag{A.47}$$

$$(\zeta_{lh} + \lambda^i)\mu(x_h^i, l) = \zeta_{hl}\mu(x_h^i, h) = \zeta_{hl}(1 - \pi - \mu(x_l^i, h)) \tag{A.48}$$

Thus,

$$\mu(x_l^i, h) = \mu(x_h^i, l) = \frac{\frac{\zeta_{hl}\zeta_{lh}}{\zeta_{hl}+\zeta_{lh}}}{\zeta_{lh} + \zeta_{hl} + \lambda^i} \tag{A.49}$$

$$\mu(x_h^i, h) = 1 - \pi - \frac{\frac{\zeta_{hl}\zeta_{lh}}{\zeta_{hl}+\zeta_{lh}}}{\zeta_{lh} + \zeta_{hl} + \lambda^i} \quad \text{and} \quad \mu(x_l^i, l) = \pi - \frac{\frac{\zeta_{hl}\zeta_{lh}}{\zeta_{hl}+\zeta_{lh}}}{\zeta_{lh} + \zeta_{hl} + \lambda^i} \tag{A.50}$$

From (A.49)-(A.50), $\mu(x_h^i, h) + \mu(x_h^i, l) = 1 - \pi$ and $\mu(x_l^i, h) + \mu(x_l^i, l) = \pi$. Let X^i denote the total asset i holdings by investors. It is thus given by

$$X^i = [\mu(x_h^i, h) + \mu(x_h^i, l)]x_h^i + [\mu(x_l^i, h) + \mu(x_l^i, l)]x_l^i = (1 - \pi)x_h^i + \pi x_l^i \tag{A.51}$$

From (3.13) and (3.14),

$$X^i = \frac{\bar{D} - (r + \pi\epsilon + \Delta^i)\bar{P}^i}{\gamma\sigma^2} \tag{A.52}$$

where

$$\begin{aligned}\Delta^i &= \frac{(1-\pi)\zeta_{hl}\epsilon}{r + \epsilon + \zeta_{lh} + \zeta_{hl} + \lambda^i\theta_l} - \frac{\pi\zeta_{lh}\epsilon}{r + \zeta_{lh} + \zeta_{hl} + \lambda^i\theta_h} \\ &= \frac{\frac{\zeta_{hl}\zeta_{lh}\epsilon}{\zeta_{hl}+\zeta_{lh}}[\lambda^i(\theta_h - \theta_l) - \epsilon]}{(r + \epsilon + \zeta_{lh} + \zeta_{hl} + \lambda^i\theta_l)(r + \zeta_{lh} + \zeta_{hl} + \lambda^i\theta_h)}\end{aligned} \tag{A.53}$$

The market clearing condition (3.16) requires that $X^i = s$. Substituting the market clearing condition into (A.52) yields the asset's inter-dealer price

$$\bar{P}^i = \frac{\bar{D} - \gamma\sigma^2 s}{r + \pi\epsilon + \Delta^i} \tag{A.54}$$

A.5 Proof of Lemma 6

$\Delta(\lambda)$ is given by (4.4). From (3.6), $\theta_h = 1 - z$ and $\theta_l = \frac{(1-z)r}{r+z\epsilon}$. Hence, $\theta_h - \theta_l = \frac{z(1-z)\epsilon}{r+z\epsilon}$. Let $k_1 \equiv r + \epsilon + \zeta_{lh} + \zeta_{hl}$ and $k_2 \equiv r + \zeta_{lh} + \zeta_{hl}$, and

$$K(\lambda) \equiv \frac{\lambda(\theta_h - \theta_l) - \epsilon}{(k_1 + \lambda\theta_l)(k_2 + \lambda\theta_h)} \tag{A.55}$$

Thus,

$$\Delta(\lambda) = \frac{\zeta_{hl}\zeta_{lh}\epsilon}{\zeta_{hl} + \zeta_{lh}} K(\lambda) \tag{A.56}$$

Differentiate $K(\lambda)$ with respect to λ,

$$\frac{\partial K(\lambda)}{\partial \lambda} = \frac{(\theta_h - \theta_l)(k_1 k_2 - \theta_l\theta_h\lambda^2) + \epsilon(k_1\theta_h + k_2\theta_l + 2\theta_l\theta_h\lambda)}{(k_1 + \lambda\theta_l)^2(k_2 + \lambda\theta_h)^2} \tag{A.57}$$

Let $\underline{\lambda}$ and $\overline{\lambda}$ denote the roots of the quadratic equation

$$(\theta_h - \theta_l)(k_1 k_2 - \theta_l\theta_h\lambda^2) + \epsilon(k_1\theta_h + k_2\theta_l + 2\theta_l\theta_h\lambda) = 0 \tag{A.58}$$

where $\underline{\lambda} < \overline{\lambda}$. Thus,

$$\underline{\lambda} = \frac{\epsilon\theta_l\theta_h - \sqrt{\epsilon^2\theta_l^2\theta_h^2 + (\theta_h - \theta_l)\theta_l\theta_h[(\theta_h - \theta_l)k_1k_2 + \epsilon(k_1\theta_h + k_2\theta_l)]}}{(\theta_h - \theta_l)\theta_l\theta_h} < 0 \tag{A.59}$$

$$\overline{\lambda} = \frac{\epsilon\theta_l\theta_h + \sqrt{\epsilon^2\theta_l^2\theta_h^2 + (\theta_h - \theta_l)\theta_l\theta_h[(\theta_h - \theta_l)k_1k_2 + \epsilon(k_1\theta_h + k_2\theta_l)]}}{(\theta_h - \theta_l)\theta_l\theta_h} > 0 \tag{A.60}$$

Because $\lambda > 0$, it is immediate that $\partial K(\lambda)/\partial\lambda > 0$ if $\lambda < \bar{\lambda}$, and $\partial K(\lambda)/\partial\lambda < 0$ if $\lambda > \bar{\lambda}$. From (A.56), $\partial\Delta(\lambda)/\partial\lambda > 0$ if $\lambda < \bar{\lambda}$, and $\partial\Delta(\lambda)/\partial\lambda < 0$ if $\lambda > \bar{\lambda}$.

A.6 Proof of Proposition 8

Note that for $\xi \in \{h, l\}$,

$$
\begin{aligned}
&V_\xi\big(x^1, \ldots, x^i + q_\xi^i(\mathbf{x}), \ldots, x^N\big) - V_\xi(\mathbf{x}) \\
&\qquad\qquad = V_{\xi i}\big(x^1, \ldots, x^i + q_\xi^i(\mathbf{x}), \ldots, x^N\big) q_\xi^i(\mathbf{x}) - L_\xi^i q_\xi^i(\mathbf{x})^2
\end{aligned}
\tag{A.61}
$$

From (3.4),

$$
V_\xi\big(x^1, \ldots, x^i + q_\xi^i(\mathbf{x}), \ldots, x^N\big) - V_\xi(\mathbf{x}) = \bar{P}^i q_\xi^i(\mathbf{x}) - L_\xi^i q_\xi^i(\mathbf{x})^2 \tag{A.62}
$$

Substitute into (3.5), and divide both sides by $q_\xi^i(\mathbf{x})$,

$$
P_\xi^i(\mathbf{x}) = \bar{P}^i - (1 - \theta_\xi) L_\xi^i q_\xi^i(\mathbf{x}) \tag{A.63}
$$

The transaction costs incurred by selling investors and buying investors are respectively

$$
Cost^{i,sell} = \frac{\bar{P}^i - P_l^i(\mathbf{x})|_{x^i = x_h^i}}{\bar{P}^i} = -\frac{1}{\bar{P}^i}(1 - \theta_l) L_l^i (x_h^i - x_l^i) \tag{A.64}
$$

$$
Cost^{i,buy} = \frac{P_h^i(\mathbf{x})|_{x^i = x_l^i} - \bar{P}^i}{\bar{P}^i} = -\frac{1}{\bar{P}^i}(1 - \theta_h) L_h^i (x_h^i - x_l^i) \tag{A.65}
$$

From (A.38) and (A.39),

$$
x_h^i - x_l^i = \frac{1 - \frac{\zeta_{lh}}{r + \zeta_{lh} + \zeta_{hl} + \lambda^i \theta_h} - \frac{\zeta_{hl}}{r + \epsilon + \zeta_{lh} + \zeta_{hl} + \lambda^i \theta_l}}{\gamma \sigma^2} \epsilon \bar{P}^i \tag{A.66}
$$

Moreover, L_h^i and L_l^i are given by (A.34) and (A.35) respectively. Substitute (A.66) and (A.35) into (A.64),

$$
\begin{aligned}
Cost^{i,sell} &= \frac{(1 - \theta_l)\epsilon}{2} \frac{r + \zeta_{hl} + \lambda^i \theta_h - \frac{(r + \zeta_{lh} + \zeta_{hl} + \lambda^i \theta_h)\zeta_{hl}}{r + \epsilon + \zeta_{lh} + \zeta_{hl} + \lambda^i \theta_l}}{(r + \zeta_{hl} + \lambda^i \theta_h)(r + \epsilon + \zeta_{lh} + \lambda^i \theta_l) - \zeta_{hl}\zeta_{lh}} \\
&= \frac{(1 - \theta_l)\epsilon}{2(r + \epsilon + \zeta_{lh} + \zeta_{hl} + \lambda^i \theta_l)}
\end{aligned}
\tag{A.67}
$$

Similarly,

$$Cost^{i,buy} = \frac{(1-\theta_h)\epsilon}{2(r+\zeta_{lh}+\zeta_{hl}+\lambda^i\theta_h)} \tag{A.68}$$

B Walrasian Benchmark: Proofs

In the Walrasian benchmark case, all assets are traded in frictionless Walrasian markets. The markets are not intermediated, and there is no role for dealers. Let $\bar{P}^{i,w}$ denote the clearing price of asset i. The state space can be decomposed into an inaction region and an action region with respect to the assets. In the inaction region, the Hamilton-Jacobi-Bellman (HJB) equation for a type h investor with asset holdings $\mathbf{x}$ is given by

$$rV_h^w(\mathbf{x}) = \sum_{i=1}^{N}\bar{D}x^i - \frac{\gamma}{2}\sum_{i=1}^{N}\sigma^2(x^i)^2 + \zeta_{hl}\Big[V_l^w(\mathbf{x}) - V_h^w(\mathbf{x})\Big] \tag{B.1}$$

where $V_\xi^w(\mathbf{x})$ is the expected value of a type $\xi \in \{h,l\}$ investor with asset holdings $\mathbf{x} \equiv (x^1,\ldots,x^N)$ in the benchmark case. In the action region, the value function satisfies

$$V_h^w(\mathbf{x}) = \max_{\{\bar{x}^i\}_i} V_h^w(\bar{x}^1,\ldots,\bar{x}^N) - \sum_{i=1}^{N}\bar{P}^{i,w}(\bar{x}^i - x^i) \tag{B.2}$$

Analogously, the HJB equation for a type l investor with asset holdings $\mathbf{x}$ in the inaction region is given by

$$(r+\epsilon)V_l^w(\mathbf{x}) = \sum_{i=1}^{N}\bar{D}x^i - \frac{\gamma}{2}\sum_{i=1}^{N}\sigma^2(x^i)^2 + \zeta_{lh}\Big[V_h^w(\mathbf{x}) - V_l^w(\mathbf{x})\Big] \tag{B.3}$$

and the HJB equation for the type l investor in the action region is

$$V_l^w(\mathbf{x}) = \max_{\{\bar{x}^i\}_i} V_l^w(\bar{x}^1,\ldots,\bar{x}^N) - \sum_{i=1}^{N}\bar{P}^{i,w}(\bar{x}^i - x^i) \tag{B.4}$$

Because the investor has continuous access to trading in all asset markets, the investor dynamically adjusts their asset holdings to re-optimize. Combining (B.1) and (B.4), the HJB equation for a type h

investor with asset holdings $\mathbf{x}$ can be written as

$$\begin{aligned} rV_h^w(\mathbf{x}) =& \sum_{i=1}^{N} \bar{D}x^i - \frac{\gamma}{2}\sum_{i=1}^{N}\sigma^2(x^i)^2 \\ &+ \zeta_{hl}\Big[\max_{\{\bar{x}^i\}_i} V_l^w(\bar{x}^1,\ldots,\bar{x}^N) - V_h^w(\mathbf{x}) - \sum_{i=1}^{N}\bar{P}^{i,w}(\bar{x}^i - x^i)\Big] \end{aligned} \tag{B.5}$$

Similarly, the HJB equation for a type l investor with asset holdings $\mathbf{x}$ is

$$\begin{aligned} (r+\epsilon)V_l^w(\mathbf{x}) =& \sum_{i=1}^{N} \bar{D}x^i - \frac{\gamma}{2}\sum_{i=1}^{N}\sigma^2(x^i)^2 \\ &+ \zeta_{lh}\Big[\max_{\{\bar{x}^i\}_i} V_h^w(\bar{x}^1,\ldots,\bar{x}^N) - V_l^w(\mathbf{x}) - \sum_{i=1}^{N}\bar{P}^{i,w}(\bar{x}^i - x^i)\Big] \end{aligned} \tag{B.6}$$

Using the first-order condition for asset position and the envelope theorem, $\forall\, i \in \{1, 2, \ldots, N\}$,

$$V_{hi}^w(\mathbf{x}) = \bar{P}^{i,w} \tag{B.7}$$

$$V_{li}^w(\mathbf{x}) = \bar{P}^{i,w} \tag{B.8}$$

and

$$rV_{hi}^w(\mathbf{x}) = \bar{D} - \gamma\sigma^2 x^i + \zeta_{hl}\Big[\bar{P}^{i,w} - V_{hi}^w(\mathbf{x})\Big] \tag{B.9}$$

$$(r+\epsilon)V_{li}^w(\mathbf{x}) = \bar{D} - \gamma\sigma^2 x^i + \zeta_{lh}\Big[\bar{P}^{i,w} - V_{li}^w(\mathbf{x})\Big] \tag{B.10}$$

Thus, the optimal asset i demand by type h and type l investors are respectively

$$x_h^{i,w} = \frac{\bar{D} - r\bar{P}^{i,w}}{\gamma\sigma^2} \tag{B.11}$$

$$x_l^{i,w} = \frac{\bar{D} - (r+\epsilon)\bar{P}^{i,w}}{\gamma\sigma^2} \tag{B.12}$$

I now derive the steady-state proportions of type h and type l investors. Let π_t denote the fraction of impatient (type l) investors. The rate of change in π_t follows

$$\dot{\pi}_t = -\zeta_{lh}\pi_t + \zeta_{hl}(1-\pi_t) \tag{B.13}$$

The solution to the above ODE is

$$\pi_t = \pi_0 e^{-(\zeta_{hl}+\zeta_{lh})t} + \frac{\zeta_{hl}}{\zeta_{hl}+\zeta_{lh}}\left[1 - e^{-(\zeta_{hl}+\zeta_{lh})t}\right] \tag{B.14}$$

As $t \to \infty$,

$$\pi_t \to \pi = \frac{\zeta_{hl}}{\zeta_{hl}+\zeta_{lh}} \tag{B.15}$$

π is the steady-state measure of type l investors. The steady-state measure of type h investors can be similarly derived as

$$1 - \pi = \frac{\zeta_{lh}}{\zeta_{hl}+\zeta_{lh}} \tag{B.16}$$

Market clearing in the asset i market requires that

$$(1-\pi)x_h^{i,w} + \pi x_l^{i,w} = \frac{\bar{D} - (r+\pi\epsilon)\bar{P}^{i,w}}{\gamma\sigma^2} = s \tag{B.17}$$

Thus, the equilibrium price is given by

$$\bar{P}^{i,w} = \frac{\bar{D} - \gamma\sigma^2 s}{r + \pi\epsilon} \tag{B.18}$$